

AD/A-003 518

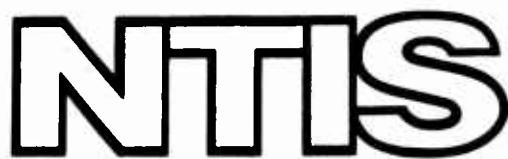
LANGUAGE IDENTIFICATION BY STATISTICAL
ANALYSIS

Morton David Rau

Naval Postgraduate School
Monterey, California

September 1974

DISTRIBUTED BY:



National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD/A002 518

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Language Identification by Statistical Analysis		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis September 1974
7. AUTHOR(s) Morton David Rau		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE September 1974
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Postgraduate School Monterey, California 93940		13. NUMBER OF PAGES 128
15. SECURITY CLASS. (of this report) Unclassified		
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE		
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Reproduced by NATIONAL TECHNICAL INFORMATION SERVICE US Department of Commerce Springfield, VA 22151		
19. KEY WORDS (Continue on reverse side if necessary and identify w/ block number) Language Identification, Linguistic Analysis, Linguistics, Statistical Linguistics		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) An analysis was conducted of English and Spanish text. The statistical analysis determined the independent probability of letters and the joint probability of various letter combinations for large samples of each language. Various methods were tested in an attempt to utilize these characteristics to identify the language of a short sample text. By use of the joint		

DD FORM 1 JAN 73 1473
(Page 1)EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) *1/28*

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

probability of various vowel-consonant relationships and the Kolmogorov-Smirnov Goodness of Fit Test an identification system was defined that provided a significance level of .0077 for a sample of 107 letters (approximately 21 words). Investigation also showed that the space rate or the interword structure in each language contains a measure of intelligence and was useful in identification.

DD Form 1473 (BACK)
1 Jan 73
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Language Identification

by

Statistical Analysis

by

Morton David Rau
Lieutenant Commander, United States Navy
B.S., Tulane University, 1962

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
September 1974

Author

Morton D Rau

Approved by:

R. A. Beitzman
Co-Advisor

V. Michael Powers

Co-Advisor

D. H. Johnson
Chairman, Department of Operations Research/
Administrative Sciences

J. R. Courtney

Academic Dean

ABSTRACT

An analysis was conducted of English and Spanish text. The statistical analysis determined the independent probability of letters and the joint probability of various letter combinations for large samples of each language.

Various methods were tested in an attempt to utilize these characteristics to identify the language of a short sample text. By use of the joint probability of various vowel-consonant relationships and the Kolmogorov-Smirnov Goodness of Fit Test an identification system was defined that provided a significance level of .0077 for a sample of 107 letters (approximately 21 words). Investigation also showed that the space rate or the interword structure in each language contains a measure of intelligence and was useful in identification.

TABLE OF CONTENTS

I.	INTRODUCTION -----	8
II.	BACKGROUND -----	10
A.	ANALYTICAL CONSIDERATIONS OF THE MODERN LATIN ALPHABET -----	10
B.	YULE'S CHARACTERISTIC (K) -----	11
C.	KOLMOGOROV-SMIRNOV TEST OF GOODNESS OF FIT -----	13
III.	EXPERIMENTS -----	15
A.	GENERAL ANALYSIS CONSIDERATIONS -----	15
B.	GENERAL PROGRAMMING CONSIDERATIONS -----	15
C.	POPULATION PARAMETERS -----	16
D.	TESTING PROCEDURES -----	18
1.	Testing by Yule's K -----	19
2.	Testing by Kolmogorov-Smirnov Goodness of Fit Test -----	20
E.	RESULTS -----	21
IV.	CONCLUSIONS -----	24
APPENDIX A.	English Population Values (V-1) -----	26
APPENDIX B.	Spanish Population Values (V-1) -----	32
APPENDIX C.	English Population Values (V-2) -----	38
APPENDIX D.	Spanish Population Values (V-2) -----	44
APPENDIX E.	Significance Levels for Kolmogorov-Smirnov Test -----	50
APPENDIX F.	Final Test Sources -----	51
APPENDIX G.	Population Analysis Program (V-1) -----	57
APPENDIX H.	Population Analysis Program (V-2) -----	71

APPENDIX I. Language Identification Program (V-1) -----	87
APPENDIX J. Language Identification Program (V-2) -----	106
LIST OF REFERENCES -----	126
INITIAL DISTRIBUTION LIST -----	128

LIST OF TABLES

Table	Title	Page
I.	Proportion of Correct Identifications -----	22
II.	English Joint Probability Table for Vowel-Consonant -----	26
III.	English (V-1) Independent Probability of X_n -----	27
IV.	English (V-1) Joint Probability Table for X_n, X_{n+1} -----	28
V.	English (V-1) Joint Probability Table for $X_n, -, X_{n+2}$ -----	30
VI.	Spanish Joint Probability Table for Vowel-Consonant -----	32
VII.	Spanish (V-1) Independent Probability of X_n -----	33
VIII.	Spanish (V-1) Joint Probability Table for X_n, X_{n+1} -----	34
IX.	Spanish (V-1) Joint Probability Table for $X_n, -, X_{n+2}$ -----	36
X.	English (V-2) Independent Probability of X_n -----	39
XI.	English (V-2) Joint Probability Table for X_n, X_{n+1} -----	40
XII.	English (V-2) Joint Probability Table for $X_n, -, X_{n+2}$ -----	42
XIII.	Spanish (V-2) Independent Probability of X_n -----	45
XIV.	Spanish (V-2) Joint Probability Table for X_n, X_{n+1} -----	46
XV.	Spanish (V-2) Joint Probability Table for $X_n, -, X_{n+2}$ -----	48
XVI.	Significance Levels for Kolmogorov-Smirnov Test -----	50

I. INTRODUCTION

In today's world of ever increasing written collections it requires a certain level of expertise to properly identify and file the material according to its language. In addition to libraries this requirement exists within large government agencies such as the State Department and in various intelligence collection agencies. Although it would be expedient for the staff of these repositories to recognize and understand every language used it does not seem to be a practical requirement. It would at least seem desirable that these persons be provided with a simple method for identifying the language even if they were unable to read the words. An approach to this problem would be to have the clerk utilize a computer terminal to enter a short sample of the text which would be analyzed by the computer and then display the identified language on the terminal. The question that arises is how would the computer determine what language the sample represented. This was the problem approached in this research.

There have been many works published on the subject of mathematical linguistics most of which seem to concern themselves with identifying the author of various literary works or style analysis of individuals or regions. One investigator, Yukio Nakamura, has developed a language identification method for some twenty-five languages based upon the probability of occurrence of various specific characters and words in each language [Ref. 1].

Since language is basically a code used to convey information it would seem that each should have a particular characteristic that would

allow identification based on a statistical pattern. It was the intent of this research to design a set of computer programs which would analyze character frequency and order of occurrence in languages which are written in the Modern Latin Alphabet and then to utilize this data to identify an unknown sample.

II. BACKGROUND

A. ANALYTICAL CONSIDERATIONS OF THE MODERN LATIN ALPHABET

The Modern Latin Alphabet consists of the basic twenty-six characters, additional special characters and characters combined with diacritical signs. Very few languages utilize only the basic twenty-six characters. English is one which relies upon the basic twenty-six, except for loan words and less fashionable use of diacritics, such as 'coördinate'.

Many languages utilize additional characters which can be divided into three classifications:

1. Single characters, e.g., ß in German.

2. Joint characters, e.g., œ in French.

3. Combined characters, e.g., ch in German or Spanish. In some languages these combinations are looked upon as individual letters and have their appropriate place in the alphabet. For the purposes of this analysis they have been considered to be two individual letters.

Diacritical marks are generally used to indicate a difference in pronunciation of a letter but in some languages the marked letter is considered to be a distinct letter, such as the ü in German. In basic linguistics they are classified as modified letters and accordingly for the purposes of this study were treated as forms of the basic letter.

Having defined the criteria for the alphabet to be analyzed the question which arose was to what extent should the interrelationship of letters be considered. It was decided that limiting the interrelationship considerations to three adjacent letters would provide sufficient

indication of trends in letter relationships without broaching too greatly into the influences of word structure. Given any three adjacent characters (x_n , x_{n+1} , x_{n+2}) it was decided to evaluate the frequencies of the following events:

1. Independent occurrence of each character, x_n .
2. Joint occurrence of each pair, x_n , x_{n+1} .
3. x_n , -, x_{n+2} : the joint occurrence of a given letter x_{n+2} as the second letter following a given x_n .
4. Vowel-Consonant relationship in the form of the occurrence of:
 - a. vowel following vowel;
 - b. consonant following vowel;
 - c. vowel following consonant;
 - d. consonant following consonant.

Such characteristics as the occurrence of x_n , x_{n+1} generate 676 (26 x 26) combinations for the English Language. This has to be compared to at least the same number from any given sample. An extensive review of works on statistical linguistic analysis and nonparametric statistics resulted in the selection of two tests as a possible means of comparing a sample with a language population. These were a characteristic 'K' devised by the British statistician G. Udny Yule and the Kolmogorov-Smirnov Test of Goodness of Fit, which are discussed below.

B. YULE'S CHARACTERISTIC (K)

G. Udny Yule was involved in conducting statistical analysis of accident distribution among various people when faced with the problem of measuring the liability of classes of people independent of the period of exposure to risk or the total number of accidents met by the whole

group at risk. As a solution to this problem he developed an equation which yielded what he termed the 'characteristic' which was nothing more than a measure of the whole group's liability. By comparing accident rate or liability to repetitiveness of words in a literary work he proposed the use of this characteristic as a measure of repetitiveness in literature. The characteristic is expressed simply as a numerical value, such as 10.6, or 97.5, or 874.3. Since the analysis of each language and sample was to measure frequency of occurrence it was felt that this measure could be said to be analogous to repetitiveness as defined by Yule.

In general, Yule's approach was to determine for a given literary work how many common nouns occurred (were used) once, how many occurred twice, thrice, etc. He then performed the following calculations to obtain the characteristic (K).

X = occurrence category, i.e., X times occurring

f_X = frequency of X , i.e., the number of words occurring X times

$$S_1 = \sum_{x=1}^n f_x x$$

$$S_2 = \sum_{x=1}^n f_x x^2$$

$$K = 10,000 \frac{S_2 - S_1}{(S_1)^2}$$

The number 10,000 was introduced simply to avoid the inconvenience of handling very small numbers [Ref. 2].

Adaptation of this measurement to the $m \times n$ matrix generated by determination of letter occurrences was as follows:

$$S_1 = \sum_{x=1}^n f_x X = f_1(1) + f_2(2) + \dots + f_n(n)$$

$$S_2 = \sum_{x=1}^n f_x X^2 = f_1(1)^2 + f_2(2)^2 + \dots + f_n(n)^2$$

Instead of X being used to relate the frequency of a word it was equated to the frequency of a specific letter combination.

\therefore Let: $X = Y_{m,n}$ = frequency of occurrence of the mn letter combination

$$f_x = f_{Y_{m,n}} = 1$$

$$\therefore S_1 = 1(Y_{1,1}) + 1(Y_{1,2}) + \dots + 1(Y_{m,n})$$

$$S_1 = \sum_{m=1}^m \sum_{n=1}^n Y_{m,n}$$

Similarly:

$$S_2 = \sum_{m=1}^m \sum_{n=1}^n (Y_{m,n})^2$$

This adaptation yields a single value or measure of repetitiveness in a specific category for each language population and sample.

C. KOLMOGOROV-SMIRNOV TEST OF GOODNESS OF FIT

Let X be a random variable with the continuous probability distribution function $f(x) = \text{Prob}[X \leq x]$ and let x_1, x_2, \dots, x_n be a sample of independent variables of size n for X , ordered so that $x_1 \leq x_2 \leq \dots \leq x_n$ which determines the empirical distribution function

$$f_n(x) = \begin{cases} 0 & \text{for } x < x_1 \\ (k/n) & \text{for } x_k \leq x \leq x_{k+1}, \quad k = 1, 2, \dots, n-1, \\ 1 & \text{for } x_n \leq x \end{cases}$$

It should be expected that for n large, $f_n(x)$ would very likely approach $f(x)$. Kolmogorov defined the statistic

$$D_n = \text{least upper bound of } |f(x) - f_n(x)|$$

which measures the greatest absolute difference between $f(x)$ and $f_n(x)$.

He further showed that it has the following properties which assists in determining how closely $f_n(x)$ represents the population $f(x)$:

1. the probability distribution of D_n depends on n but is independent of $f(x)$, (the maximum D_n is a random variable)
2. the probability distribution of D_n is given by the relationship

$$\lim_{n \rightarrow \infty} \text{Prob} \left[D_n < \frac{z}{\sqrt{n}} \right] = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 z^2} = L(z)$$

[Refs. 3, 4, 5, and 6].

The function $L(z)$ has been tabulated by N. Smirnov [Ref. 7]. From the equation $D_n \sqrt{n} = z$ a value of z is obtained to enter the tables and thereby determine $\text{Prob} \left[D_n < \frac{z}{\sqrt{n}} \right]$ or find the significance level of the fit as $1 - L(z)$. These values have been computed and are tabulated in Appendix E.

Prior to continuing some comment should be made as to the selection of the Kolmogorov-Smirnov Test over the Chi-Squared Test. The Chi-Squared test requires a large sample size, because n at finite values is not distribution-free and only approaches this condition as n approaches infinity [Refs. 4 and 8].

III. EXPERIMENTS

A. GENERAL ANALYSIS CONSIDERATIONS

In considering the alphabetic structure of a language there is no doubt as to the fact that the letters carry a certain amount of intelligence but it was not certain as to what effect if any was played by the space, other than to separate words. In order to determine the effect of the space as a character in identifying a language two analytical approaches were taken in evaluating the characteristics described in section II. A. The first (version one, Appendix G) was to conduct the analysis considering spaces as separators, not characters. All characters not described as alphabetic, such as the period, comma, etc., were considered to be spaces and signified only the termination of a word. Joint frequencies such as that of x_n, x_{n+1} were counted only when the pair occurred within the same word. This resulted in the measurement of the desired characteristics exclusively within words, without measuring the interword relationships. The second approach (version two, Appendix H) was to consider the space and nonalphabetic characters between words as a space character. This resulted in the measurement of such characteristics as frequency of first letter occurrence, frequency of last letter occurrences, and the interword relationship of last letter to first letter.

B. GENERAL PROGRAMMING CONSIDERATIONS

All computer programs were written in American National Standard COBOL.

Input language samples were fed to the programs via standard eighty column punch cards in the following format:

<u>Card Column</u>	<u>Content</u>
1, 2	Blank
3 - 72	Text
73	Blank unless the last word on the previous card could not be completed in column 72, then a '--' was placed in column 73 to signify continuation on the next card.
74 - 80	Sequence Number

Only alphabetic text with normal punctuation was used for input text. Numeric characters appearing in the various text material were converted to their alphabetic equivalent. To allow for the compilation of data from languages which utilize special characters the keypunch characters 1, 2, 3, 4, and 5 were allocated to those classified as vowels while 6 and 7 were for consonants. This conversion, if required, had to be performed by the operator at the time of entry or during data preparation.

In those programs that were written to consider the space as a character of the alphabet, all interword spaces and punctuation marks were compressed to one space.

C. POPULATION PARAMETERS

The actual population in the case of written text is potentially infinite. Due to limited resources it was decided to test the theory using representative samples of two languages as a measure of their respective populations. The source of these population samples was

chosen to be Reader's Digest as it is available in many languages and contains several short articles, by different authors, which have been selected from various areas of interest, thereby suppressing the influence of one person's or region's style on the assumed population.

The English population consisted of text material, excluding advertisements, from Reader's Digest, Vol. 104, No. 623, March 1974, pages 8, 11, 12, 31 - 90, 109 and 110. This text consists of 79,185 characters excluding spaces and 96,100 characters including spaces.

The Spanish population consisted of text material, excluding advertisements, from Selecciones Del Reader's Digest, Vol. 67, No. 402, May 1974, pages 3 - 10, 21 - 30, 37-60, and 75 -80. This text consists of 78,851 characters excluding spaces and 94,992 characters including spaces.

Both populations were analyzed for the characteristics described in section II. A. (with and without space consideration) and the following population means were obtained:

1. Yule's 'K' based on the independent occurrence of each character, X_n ;
2. Yule's 'K' based on the joint occurrence of X_n, X_{n+1} ;
3. Yule's 'K' based on the joint occurrence of $X_n, -, X_{n+2}$;
4. Probability table for the independent occurrence of X_n ;
5. Joint probability table for the occurrence of X_n, X_{n+1} ;
6. Joint probability table for the occurrence of $X_n, -, X_{n+2}$;
7. Joint probability table for vowel-consonant occurrence.

(A tabulation of these values is contained in Appendices A, B, C, and D)

Each probability table was formed into a cumulative relative frequency

curve, the points of which were transferred to punched cards, for later use in testing samples by the Kolmogorov-Smirnov Goodness of Fit Test.

In order to facilitate the use of Yule's 'K' in testing small samples it was necessary to approximate the standard deviation (σ) for each value of Yule's 'K'. The population was divided into groups of ten card samples from which a new 'K' was determined for each sample. The following calculations were then performed:

K = Yule's 'K' obtained from each sample of 10 cards (less than 700 characters).

n = number of 10 card samples contained in the population.

$$\bar{K} = \frac{\sum K}{n} = \text{mean value of Yule's 'K'}$$

$$\therefore \sigma_{10} = \sqrt{\frac{(\sum K^2 - n\bar{K})}{(n - 1)}}$$

σ_{10} was then adjusted to a population standard deviation based on one card samples by computing

$$\sigma = \sigma_{10} \sqrt{10}$$

[Ref. 8].

(Values of σ are contained in Appendices A, B, C, and D.) A histogram was plotted for the various values of Yule's 'K' and they appeared to conform to a normal distribution.

D. TESTING PROCEDURES

In constructing testing procedures there were two approaches which were considered. The first was that a small sample would be analyzed and identified as either English or Spanish. The second approach was to analyze a small sample and then report it as either English, Spanish, or undecisive based upon some predefined constraints. Since it was

intended, if the theory worked, that the process could be expanded to cover all languages written in the Modern Latin Alphabet it was felt that the first approach was more realistic (useful).

1. Testing by Yule's K

Since the incremental population samples of ten cards appeared to conform to the normal distribution it was assumed that for all languages the value of K obtained from small samples would conform to the normal distribution. For each sample tested three values of K were obtained, one for each of the three K tests. Each value of K was then compared to its respective population mean, μ , obtained from the representative population analysis, in each language in order to determine the appropriate normal deviate, z, by

$$z = \frac{K - \mu}{\sigma / \sqrt{n}}$$

where n was the number of data cards in that particular sample. For each of the three tests the z for English was compared with the z for that test in Spanish. The smallest absolute value of z identified the language or that language to which the sample had the highest probability of belonging. The smallest z from each test was then used to compute the significance level, α , as a measure of how closely K and μ matched based on the probability of a normal distribution. For absolute values of z less than 4.0, α was computed as follows:

$$\begin{aligned} x &= |z| \\ \alpha &= 1 - 2x(.39894,22804,014) \left[1 - \frac{1}{6}x^2 + \frac{P}{6.66666,66666,67} x^2 \right. \\ &\quad \left. - \frac{P}{8.4} + \frac{P}{10.28571,42857,1} x^2 - \frac{P}{12.22222,22222,2} x^2 \right. \\ &\quad \left. + \frac{P}{14.18181,81818,2} x^2 - \frac{P}{16.15384,61538,5} x^2 \right] \end{aligned}$$

$$\begin{aligned}
& + \frac{P}{18.13333,33333,3} x^2 - \frac{P}{20.11764,70588,2} x^2 \\
& + \frac{P}{22.10526,31578,9} x^2 - \frac{P}{24.09523,80952,4} x^2 \\
& + \frac{P}{26.08695,65217,4} x^2 - \frac{P}{28.08000,0000,0} x^2 \\
& + \frac{P}{30.07407,40740,7} x^2 - \frac{P}{32.06896,55172,4} x^2 \\
& + \frac{P}{34.06451,61290,3} x^2 - \frac{P}{36.06060,60606,1} x^2 \\
& + \frac{P}{38.05714,28571,4} x^2 - \frac{P}{40.05405,40540,5} x^2 \\
& + \frac{P}{42.05128,20512,8} x^2 - \frac{P}{44.04878,04878,0} x^2 \\
& + \frac{P}{46.04651,16279,1} x^2 - \frac{P}{48.04444,44444,4} x^2 \\
& + \frac{P}{50.04255,3191} x^2 - \frac{P}{52.04081,6326} x^2 \\
& + \frac{P}{54.03921,5686} x^2 - \frac{P}{56.03773,5849} x^2 \\
& + \frac{P}{58.03636} x^2 - \frac{P}{60.03509} x^2 + \frac{P}{62.03} x^2
\end{aligned}
]$$

where P = the absolute value of the preceding term [Ref. 9]. For absolute values of z greater than or equal to 4.0, α was reported as being less than 0.00006 [Ref. 8].

2. Testing by the Kolmogorov-Smirnov Goodness of Fit Test

For each sample tested four cumulative relative frequency tables were computed, one from each of the probability tables described in section III. C. These tables represented points, $f_n(x)$, on the cumulative relative frequency (CRF) curve for each particular sample. The CRF curve points for the sample were compared against the estimate of the respective curve points, $f(x)$, along the ordinate, for the Spanish and English populations. From each comparison the value of

D_n = least upper bound of $|f(x) - f_n(x)|$ was obtained. For each of the four tests the value of D_n from the comparison with Spanish was compared to that D_n from the comparison with English. The sample was judged to be of that language which resulted in the smallest D_n . The smallest value of D_n from each test was then used to compute the significance level as a measure of how closely the sample and population CRF curves matched. D_n was then multiplied by the square root of the number of data cards contained in the sample to obtain z . For values of z greater than or equal to 0.28 and less than or equal to 1.82, the significance level was obtained from the table contained in Appendix E. For values of z less than 0.28, α was reported as greater than 0.999999. For values of z greater than 1.82, α was reported as less than 0.002645.

E. RESULTS

In an attempt to construct an unbiased set of samples for final testing, an assortment of short stories, novels, poems, reports, and speeches by one hundred and thirty different authors was selected from several anthologies (Appendix F). The samples within this group were equally split between English and Spanish and covered a period from the twelfth century through the present day. Each sample was taken from the beginning of the particular work and was of such length to fill the prescribed text field of ten punch cards.

Each sample was analyzed four times: once utilizing all ten data cards (534 characters); once utilizing the first five data cards (267 characters); once utilizing the first two data cards (107 characters); and once utilizing only the first data card (53 characters).

Furthermore, each of these were processed once under Language-ID-Version-One (Appendix I) which used the space only as a separator and once under Language-ID-Version-Two (Appendix J) which used the space as a character.

In addition to the seven separate tests described in section III. C., three other tests which were evaluated were:

1. a majority vote of all seven tests;
2. a majority vote of the three tests based on Yule's 'K';
3. a majority vote of the four tests based on the Kolmogorov-Smirnov (K-S) test of goodness of fit.

The following tabulation presents the proportion of correct identifications achieved by each of the ten tests under Language-ID-Version-One (V-1) and Language-ID-Version-Two (V-2).

TABLE I
Proportion of Correct Identifications

No. of Cards	K for X_n		K for X_n, X_{n+1}	
	V-1	V-2	V-1	V-2
10	.9538	.6615	.6000	.6231
5	.9000	.6462	.5769	.6384
2	.7615	.5615	.5846	.6385
1	.7077	.5308	.5154	.5923

	K for X_n, \dots, X_{n+2}		K-S for Vowel-Cons.	
	V-1	V-2	V-1	V-2
10	.5308	.5846	1.0000	
5	.5077	.5923	1.0000	
2	.4615	.5461	.9923	
1	.5000	.5231	.8770	

	K-S for X_n		K-S for X_n, X_{n+1}	
	V-1	V-2	V-1	V-2
10	1.0000	1.0000	.9923	1.0000
5	.9769	.9769	.9230	.9846
2	.9348	.9348	.8539	.9615
1	.8770	.8693	.7230	.8539

No. of Cards	K-S for X_n, \dots, X_{n+2}		Majority Vote	
	V-1	V-2	V-1	V-2
10	.9307	1.0000	.9846	1.0000
5	.8770	.9769	.9615	1.0000
2	.8154	.9308	.9230	.9615
1	.7521	.8153	.8923	.8923
Majority of K				
V-1	V-2	Majority of K-S		
		V-1	V-2	
10	.6385	.6154	.9923	1.0000
5	.6615	.6118	.9461	.9769
2	.6308	.5770	.8923	.9461
1	.5846	.5384	.7846	.8385

IV. CONCLUSIONS

The following observations have been made based upon the results presented in section III. E.

1. Yule's characteristic 'K' by itself was not a reliable measure for identifying the language based on a small sample. Two factors could have affected this measure: first was the close proximity of the mean values when viewed in light of the standard deviation; secondly was that with only twenty-six letters being considered there was too great a chance of repetitiveness which in fact was what Yule was measuring.

2. It was observed that a higher proportion of correct identifications occurred when the sample was analyzed using the space as a character. Accordingly it was concluded that the space rate or the interword structure in each language does contain a measure of intelligence and was useful in identification.

3. The use of the Kolmogorov-Smirnov Goodness of Fit Test when applied to vowel-consonant relationships proved to be the most reliable single test with a significance level of 0.0077 for a sample of approximately 107 characters.

4. No majority vote test was as reliable as the vowel-consonant test but the test based on a majority of the seven individual tests did prove to be the second most reliable test. In view of the fact that this test incorporated the results of three low-reliability Yule's 'K' tests it is subject to some doubt as to how reliable it would prove if the program were expanded to include more than two languages.

Although the experiments using only two languages could not produce conclusive evidence, this research tends to support the belief that each language does possess some statistical characteristic which would allow identification using a short sample of text. Expansion of this research to include additional languages would be useful in determining the reliability of this type of identification system when used with short samples of text.

APPENDIX A. ENGLISH POPULATION VALUES (V-1)

1. Yule's 'K' based on the independent occurrence of each character

x_n :

Mean value 645.079

Standard Deviation 81.479

2. Yule's 'K' based on the joint occurrence of x_n, x_{n+1} :

Mean value 87.168

Standard Deviation 40.207

3. Yule's 'K' based on the joint occurrence of $x_n, -, x_{n+2}$:

Mean value 62.083

Standard Deviation 42.821

TABLE II. ENGLISH JOINT PROBABILITY TABLE FOR VOWEL-CONSONANT

		First Letter	
		Vowel	Consonant
Second Letter	Vowel	.0512	.3568
	Consonant	.3549	.2371

TABLE III. ENGLISH (V-1) INDEPENDENT PROBABILITY OF X_n

X_n	Probability
A	.0800
E	.1214
I	.0710
O	.0774
U	.0284
B	.0138
C	.0313
D	.0418
F	.0222
G	.0217
H	.0502
J	.0017
K	.0072
L	.0445
M	.0255
N	.0731
P	.0210
Q	.0008
R	.0631
S	.0640
T	.0880
V	.0112
W	.0188
X	.0020
Y	.0191
Z	.0008

TABLE IV. ENGLISH (V-1) JOINT PROBABILITY TABLE FOR χ_n , χ_{n+1}

	χ_n																											
	A	E	I	O	U	B	C	D	F	G	H	J	K															
A	.0000	.0072	.0016	.0012	.0010	.0022	.0065	.0022	.0015	.0019	.0096	.0001	.0001															
E	.0000	.0050	.0030	.0004	.0013	.0046	.0046	.0058	.0019	.0040	.0271	.0005	.0029															
I	.0033	.0020	.0000	.0010	.0010	.0012	.0022	.0038	.0040	.0014	.0086	.0000	.0012															
O	.0000	.0007	.0062	.0025	.0001	.0021	.0085	.0038	.0058	.0016	.0065	.0006	.0001															
U	.0011	.0002	.0000	.0125	.0000	.0023	.0010	.0013	.0010	.0007	.0019	.0009	.0000															
B	.0019	.0002	.0011	.0008	.0006	.0001	.0000	.0001	.0000	.0000	.0001	.0000	.0001															
C	.0031	.0040	.0066	.0017	.0017	.0000	.0003	.0001	.0000	.0000	.0000	.0000	.0000															
D	.0039	.0133	.0035	.0017	.0017	.0007	.0000	.0000	.0005	.0000	.0000	.0000	.0000															
F	.0012	.0012	.0019	.0075	.0002	.0000	.0000	.0000	.0000	.0017	.0000	.0000	.0000															
G	.0019	.0009	.0038	.0007	.0016	.0000	.0000	.0000	.0000	.0003	.0000	.0000	.0000															
H	.0001	.0002	.0000	.0004	.0000	.0000	.0054	.0001	.0000	.0035	.0000	.0000	.0000															
J	.0001	.0000	.0000	.0001	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000															
K	.0012	.0003	.0004	.0010	.0000	.0000	.0026	.0000	.0000	.0000	.0000	.0000	.0000															
L	.0106	.0054	.0051	.0052	.0032	.0020	.0013	.0005	.0007	.0006	.0001	.0001	.0000															
M	.0031	.0032	.0037	.0052	.0012	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000															
N	.0201	.0150	.0238	.0170	.0050	.0000	.0002	.0000	.0002	.0008	.0002	.0000	.0006															
P	.0013	.0016	.0011	.0024	.0017	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000															
Q	.0000	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000															
R	.0120	.0188	.0033	.0126	.0057	.0007	.0022	.0021	.0019	.0020	.0011	.0000	.0000															
S	.0086	.0124	.0085	.0026	.0052	.0002	.0002	.0002	.0014	.0001	.0007	.0002	.0000															
T	.0155	.0045	.0093	.0039	.0047	.0001	.0030	.0001	.0018	.0002	.0024	.0000	.0000															
V	.0022	.0032	.0035	.0019	.0000	.0000	.0000	.0002	.0002	.0000	.0000	.0000	.0000															
W	.0009	.0014	.0000	.0038	.0000	.0000	.0001	.0001	.0001	.0001	.0000	.0000	.0000															
X	.0002	.0014	.0006	.0003	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000															
Y	.0029	.0014	.0000	.0002	.0001	.0015	.0004	.0007	.0001	.0005	.0003	.0000	.0000															
Z	.0002	.0001	.0006	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000															

χ_{n+1}

TABLE IV - Continued

	X_n												
	L	M	N	P	Q	R	S	T	V	W	X	Y	Z
A	.0053	.0057	.0034	.0030	.0000	.0062	.0035	.0039	.0007	.0047	.0001	.0001	.0002
E	.0095	.0081	.0090	.0048	.0000	.0174	.0080	.0128	.0102	.0047	.0001	.0012	.0005
I	.0070	.0042	.0040	.0012	.0000	.0067	.0048	.0109	.0026	.0034	.0002	.0003	.0001
O	.0035	.0037	.0037	.0035	.0000	.0077	.0033	.0110	.0007	.0031	.0000	.0022	.0001
U	.0009	.0010	.0006	.0009	.0010	.0011	.0027	.0019	.0000	.0000	.0000	.0000	.0000
B	.0000	.0008	.0000	.0000	.0000	.0002	.0002	.0001	.0000	.0000	.0000	.0000	.0000
C	.0000	.0001	.0030	.0000	.0000	.0013	.0013	.0003	.0000	.0000	.0001	.0001	.0000
D	.0031	.0000	.0151	.0000	.0000	.0021	.0001	.0000	.0000	.0000	.0000	.0000	.0000
F	.0006	.0000	.0005	.0000	.0000	.0002	.0002	.0001	.0001	.0000	.0000	.0000	.0000
G	.0000	.0000	.0114	.0000	.0000	.0019	.0000	.0000	.0000	.0000	.0000	.0000	.0000
H	.0000	.0000	.0002	.0007	.0000	.0001	.0032	.0300	.0000	.0035	.0001	.0000	.0000
J	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
K	.0002	.0000	.0008	.0000	.0000	.0011	.0005	.0000	.0000	.0000	.0000	.0000	.0000
L	.0089	.0001	.0004	.0023	.0000	.0008	.0007	.0011	.0000	.0001	.0000	.0006	.0000
M	.0002	.0007	.0004	.0001	.0000	.0012	.0011	.0002	.0000	.0000	.0000	.0001	.0000
N	.0000	.0001	.0008	.0000	.0000	.0018	.0002	.0000	.0000	.0008	.0000	.0001	.0000
P	.0004	.0025	.0001	.0016	.0000	.0004	.0016	.0016	.0001	.0000	.0005	.0000	.0000
Q	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000
R	.0002	.0001	.0001	.0046	.0000	.0011	.0000	.0040	.0000	.0002	.0000	.0003	.0000
S	.0016	.0008	.0039	.0007	.0000	.0060	.0036	.0029	.0000	.0005	.0000	.0006	.0000
T	.0008	.0000	.0100	.0007	.0000	.0036	.0102	.0017	.0000	.0000	.0007	.0001	.0000
V	.0003	.0000	.0009	.0000	.0000	.0005	.0000	.0000	.0000	.0000	.0000	.0000	.0000
W	.0001	.0000	.0001	.0000	.0000	.0001	.0004	.0018	.0000	.0000	.0000	.0000	.0000
X	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Y	.0042	.0008	.0010	.0003	.0000	.0021	.0003	.0042	.0001	.0000	.0001	.0000	.0000
Z	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000
X_{n+1}													

TABLE V. ENGLISH (V-1) JOINT PROBABILITY TABLE FOR X_n , -, X_{n+2}

		X_n												
		A	E	I	O	U	B	C	D	F	G	H	J	K
A	.0021	.0039	.0057	.0034	.0032	.0008	.0045	.0012	.0008	.0011	.0011	.0000	.0001	
E	.0175	.0150	.0168	.0151	.0045	.0025	.0042	.0034	.0030	.0023	.0015	.0001	.0005	
I	.0093	.0070	.0068	.0046	.0051	.0015	.0035	.0011	.0016	.0011	.0014	.0001	.0001	
O	.0031	.0034	.0014	.0012	.0009	.0007	.0020	.0009	.0025	.0013	.0014	.0001	.0007	
U	.0011	.0018	.0025	.0012	.0004	.0012	.0032	.0003	.0011	.C001	.0031	.0000	.0000	
B	.0003	.0007	.0002	.0004	.0006	.0002	.0004	.0001	.0001	.0002	.0002	.0000	.0000	
C	.0023	.0028	.0014	.0011	.0010	.0014	.0000	.0021	.0015	.0002	.0012	.0002	.0000	
D	.0168	.0037	.0025	.0038	.0058	.0004	.0007	.0017	.0006	.0005	.0021	.0001	.0011	
F	.0006	.0013	.0006	.0015	.0005	.0004	.0002	.0003	.0006	.0001	.0001	.0000	.0000	
G	.0023	.0027	.0125	.0037	.0009	.0007	.0005	.0002	.0001	.0000	.0008	.0000	.0000	
H	.0021	.0012	.0070	.0019	.0026	.0000	.0000	.0001	.0000	.0001	.0000	.0002	.0000	
J	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	
K	.0024	.0009	.0018	.0028	.0003	.0000	.0000	.0000	.0000	.0000	.0002	.0000	.0000	
L	.0067	.0031	.0046	.0086	.0027	.0013	.0038	.0025	.0013	.0005	.0034	.0000	.0002	
M	.0009	.0011	.0009	.0021	.0001	.0000	.0034	.0008	.0004	.0002	.0030	.0001	.0001	
N	.0026	.0041	.0098	.0051	.0012	.0015	.0085	.0032	.0011	.0018	.0067	.0002	.0015	
P	.0017	.0031	.0015	.0023	.0011	.0000	.0007	.0005	.0000	.0001	.0012	.0000	.0001	
Q	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	
R	.0021	.0055	.0005	.0053	.0008	.0016	.0024	.0023	.0078	.0020	.0062	.0004	.0004	
S	.0068	.0094	.0049	.0088	.0019	.0008	.0027	.0017	.0008	.0019	.0074	.0009	.0003	
T	.0084	.0156	.0070	.0087	.0049	.0031	.0013	.0012	.0004	.0045	.0051	.0000	.0003	
V	.0004	.0015	.0010	.0006	.0001	.0002	.0002	.0004	.0016	.0008	.0018	.0000	.0000	
W	.0001	.0003	.0002	.0001	.0001	.0001	.0001	.0006	.0003	.0000	.0012	.0001	.0001	
X	.0000	.0000	.0000	.0000	.0000	.0002	.0000	.0000	.0001	.0000	.0000	.0000	.0000	
Y	.0027	.0018	.0020	.0007	.0002	.0007	.0001	.0015	.0005	.0002	.0013	.0000	.0002	
Z	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0001	.0000	.0000	

X_{n+2}

TABLE V - Continued

	χ_n	χ_{n+2}	L	M	N	P	Q	R	S	T	V	W	X	Y	Z
A	.0042	.0016	.0023	.0017	.0005	.0047	.0053	.0069	.0001	.0011	.0014	.0000			
E	.0038	.0019	.0080	.0045	.0004	.0087	.0061	.0317	.0003	.0019	.0006	.0010	.0001		
I	.0023	.0012	.0051	.0028	.0003	.0042	.0051	.0054	.0001	.0018	.0002	.0002	.0000		
O	.0023	.0006	.0028	.0053	.0000	.0022	.0045	.0104	.0001	.0013	.0001	.0001	.0000		
U	.0007	.0005	.0011	.0004	.0000	.0022	.0024	.0011	.0000	.0011	.0000	.0023	.0000		
B	.0004	.0001	.0002	.0005	.0000	.0015	.0005	.0005	.0000	.0000	.0000	.0000	.0000		
C	.0020	.0011	.0007	.0011	.0000	.0045	.0015	.0025	.0006	.0001	.0001	.0001	.0000		
D.	.0021	.0018	.0013	.0009	.0000	.0049	.0021	.0035	.0010	.0004	.0001	.0003	.0002		
F	.0008	.0001	.0001	.0000	.0000	.0011	.0005	.0005	.0002	.0000	.0002	.0000	.0000		
G	.0014	.0007	.0005	.0001	.0000	.0017	.0004	.0006	.0006	.0001	.0000	.0000	.0000		
H	.0004	.0002	.0010	.0000	.0000	.0018	.0006	.0004	.0000	.0000	.0001	.0003	.0000		
J	.0000	.0002	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
K	.0005	.0011	.0000	.0000	.0000	.0001	.0000	.0005	.0000	.0000	.0000	.0000	.0000		
L	.0000	.0028	.0030	.0022	.0000	.0030	.0029	.0034	.0013	.0020	.0001	.0001	.0000		
M	.0011	.0002	.0009	.0000	.0000	.0041	.0025	.0026	.0002	.0002	.0001	.0000	.0000		
N	.0037	.0070	.0032	.0022	.0000	.0040	.0041	.0052	.0028	.0023	.0001	.0008	.0004		
P	.0005	.0000	.0002	.0004	.0000	.0020	.0008	.0007	.0000	.0001	.0000	.0000	.0000		
Q	.0000	.0000	.0000	.0000	.0000	.0002	.0000	.0001	.0000	.0000	.0000	.0000	.0000		
R	.0039	.0043	.0057	.0055	.0000	.0005	.0040	.0094	.0047	.0037	.0002	.0004	.0002		
S	.0042	.0035	.0042	.0016	.0000	.0059	.0015	.0018	.0015	.0031	.0002	.0002	.0000		
T	.0038	.0028	.0063	.0013	.0000	.0036	.0011	.0016	.0007	.0028	.0000	.0003	.0001		
V	.0008	.0004	.0007	.0000	.0000	.0017	.0014	.0009	.0001	.0004	.0000	.0000	.0000		
W	.0013	.0000	.0023	.0002	.0000	.0011	.0002	.0003	.0000	.0000	.0000	.0000	.0000		
X	.0000	.0001	.0003	.0000	.0000	.0000	.0008	.0003	.0000	.0000	.0000	.0000	.0000		
Y	.0018	.0006	.0022	.0009	.0000	.0025	.0010	.0017	.0000	.0011	.0004	.0000	.0000		
Z	.0002	.0001	.0004	.0000	.0000	.0001	.0001	.0001	.0000	.0000	.0000	.0000	.0000		

 χ_{n+2}

APPENDIX B. SPANISH POPULATION VALUES (V-1)

1. Yule's 'K' based on the independent occurrence of each character

x_n :

Mean value	751.447
Standard Deviation	82.328

2. Yule's 'K' based on the joint occurrence of x_n, x_{n+1} :

Mean value	104.553
Standard Deviation	38.094

3. Yule's 'K' based on the joint occurrence of $x_n, -, x_{n+2}$:

Mean value	72.680
Standard Deviation	35.210

TABLE VI. SPANISH JOINT PROBABILITY TABLE FOR VOWEL-CONSONANT

		First Letter	
		Vowel	Consonant
Second Letter	Vowel	.0616	.4469
	Consonant	.3713	.1202

TABLE VII. SPANISH (V-1) INDEPENDENT PROBABILITY OF X_n

X_n	Probability
A	.1269
E	.1311
I	.0690
O	.0929
U	.0409
B	.0137
C	.0448
D	.0470
F	.0072
G	.0111
H	.0087
J	.0041
K	.0006
L	.0568
M	.0285
N	.0747
P	.0247
Q	.0087
R	.0619
S	.0752
T	.0464
V	.0114
W	.0000
X	.0014
Y	.0089
Z	.0037

TABLE VIII. SPANISH (V-1) JOINT PROBABILITY TABLE FOR X_n , X_{n+1}

	X_n												
	A	E	I	O	U	B	C	D	F	G	H	J	K
A	.0000	.0017	.0115	.0001	.0032	.0055	.0106	.0076	.0009	.0027	.0050	.0012	.0001
E	.0003	.0003	.0108	.0000	.0146	.0013	.0049	.0278	.0015	.0015	.0012	.0013	.0001
I	.0009	.0009	.0000	.0004	.0029	.0033	.0132	.0066	.0026	.0016	.0013	.0001	.0002
O	.0003	.0011	.0104	.0003	.0002	.0013	.0133	.0119	.0015	.0024	.0028	.0019	.0001
U	.0012	.0006	.0002	.0000	.0000	.0008	.0047	.0017	.0014	.0028	.0006	.0006	.0000
B	.0066	.0006	.0017	.0020	.0008	.0000	.0C00	.0000	.0000	.0000	.0000	.0000	.0000
C	.0081	.0062	.0070	.0033	.0022	.0000	.0C11	.0000	.0000	.0000	.0000	.0000	.0000
D	.0101	.0037	.0053	.0026	.0015	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
F	.0005	.0013	.0008	.0007	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
G	.0008	.0029	.0021	.0011	.0004	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
H	.0003	.0000	.0000	.0000	.0000	.0000	.0035	.0000	.0000	.0000	.0000	.0000	.0000
J	.0009	.0014	.0004	.0004	.0003	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000
K	.0000	.0001	.0000	.0001	.0000	.0000	.0001	.0001	.0000	.0000	.0000	.0000	.0000
L	.0116	.0139	.0048	.0037	.0020	.0018	.0009	.0000	.0004	.0003	.0000	.0000	.0001
M	.0054	.0035	.0041	.0053	.0012	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000
N	.0176	.0302	.0095	.0164	.0111	.0000	.0C01	.0000	.0002	.0001	.0000	.0000	.0000
P	.0015	.0011	.0005	.0014	.0008	.0000	.0C00	.0000	.0000	.0000	.0000	.0000	.0000
Q	.0006	.0003	.0002	.0001	.0000	.0000	.0C00	.0001	.0000	.0000	.0000	.0000	.0000
R	.0173	.0162	.0028	.0102	.0031	.0026	.0C09	.0011	.0008	.0023	.0000	.0000	.0000
S	.0157	.0234	.0049	.0230	.0032	.0005	.0000	.0000	.0000	.0000	.0000	.0001	.0000
T	.0038	.0037	.0035	.0021	.0011	.0001	.0029	.0000	.0000	.0000	.0000	.0000	.0000
V	.0009	.0015	.0024	.0023	.0003	.0000	.0C00	.0001	.0000	.0000	.0000	.0000	.0000
W	.0000	.0000	.0000	.0000	.0000	.0000	.0C00	.0000	.0000	.0000	.0000	.0000	.0000
X	.0001	.0014	.0000	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Y	.0009	.0002	.0000	.0004	.0006	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Z	.0009	.0014	.0008	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

X_{n+1}

TABLE VIII. Continued

x_n	x_{n+1}	y	z
A	E	.0009	.0022
B	F	.0004	.0000
C	G	.0005	.0000
D	H	.0048	.0000
E	I	.0049	.0000
F	J	.0081	.0000
G	K	.0113	.0000
H	L	.0015	.0001
I	M	.0017	.0001
J	N	.0046	.0024
K	P	.0001	.0001
L	Q	.0000	.0000
M	R	.0000	.0000
N	S	.0000	.0000
P	T	.0000	.0000
Q	V	.0000	.0000
R	W	.0000	.0000
S	X	.0000	.0000
T	Y	.0000	.0000
V	Z	.0000	.0000

TABLE IX. SPANISH (V-1) JOINT PROBABILITY TABLE FOR X_n , -, X_{n+2}

		X_n													
		A	E	I	O	U	B	C	D	F	G	H	J	K	
		A	.0196	.0140	.0169	.0095	.0106	.0031	.0094	.0024	.0007	.0023	.0000	.0001	.0000
		E	.0128	.0113	.0056	.0084	.0030	.0047	.0052	.0022	.0020	.0003	.0001	.0000	.0000
		I	.0171	.0134	.0097	.0091	.0036	.0016	.0043	.0005	.0003	.0015	.0000	.0000	.0000
		O	.0158	.0108	.0103	.0084	.0053	.0009	.0086	.0015	.0003	.0016	.0000	.0000	.0001
		U	.0034	.0034	.0026	.0016	.0005	.0002	.0015	.0001	.0001	.0003	.0000	.0000	.0000
		B	.0011	.0005	.0003	.0012	.0003	.0003	.0010	.0007	.0002	.0001	.0025	.0002	.0000
		C	.0024	.0054	.0020	.0012	.0020	.0004	.0008	.0028	.0023	.0004	.0019	.0002	.0000
		D	.0035	.0027	.0016	.0021	.0022	.0003	.0025	.0035	.0000	.0005	.0001	.0002	.0000
		F	.0002	.0006	.0006	.0004	.0001	.0001	.0001	.0007	.0000	.0000	.0000	.0002	.0000
		G	.0021	.0005	.0007	.0011	.0009	.0000	.0002	.0001	.0002	.0000	.0001	.0001	.0000
		H	.0008	.0008	.0001	.0008	.0011	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
		J	.0000	.0001	.0003	.0003	.0001	.0005	.0000	.0005	.0000	.0001	.0002	.0000	.0000
		K	.0002	.0001	.0000	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0000
		L	.0041	.0025	.0045	.0008	.0023	.0008	.0037	.0042	.0005	.0006	.0006	.0000	.0002
		M	.0021	.0018	.0028	.0010	.0003	.0003	.0057	.0012	.0002	.0002	.0014	.0005	.0000
		N	.0014	.0020	.0160	.0003	.0031	.0017	.0112	.0035	.0017	.0027	.0011	.0007	.0000
		P	.0004	.0051	.0014	.0019	.0005	.0000	.0013	.0002	.0000	.0000	.0001	.0000	.0000
		Q	.0007	.0001	.0002	.0003	.0003	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000
		R	.0041	.0040	.0037	.0032	.0031	.0021	.0055	.0043	.0018	.0022	.0013	.0008	.0000
		S	.0021	.0028	.0047	.0021	.0028	.0008	.0050	.0086	.0006	.0009	.0016	.0010	.0000
		T	.0100	.0248	.0055	.0041	.0036	.0009	.0007	.0012	.0005	.0005	.0001	.0003	.0000
		V	.0005	.0013	.0007	.0006	.0008	.0000	.0002	.0008	.0000	.0001	.0007	.0000	.0000
		W	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
		X	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
		Y	.0000	.0001	.0000	.0000	.0000	.0001	.0001	.0000	.0000	.0005	.0000	.0001	.0000
		Z	.0009	.0006	.0009	.0000	.0001	.0001	.0002	.0000	.0001	.0003	.0000	.0000	.0000
		X_{n+2}													

TABLE IX - Continued

APPENDIX C. ENGLISH POPULATION VALUES (V-2)

1. Yule's 'K' based on the independent occurrence of each character

x_n :

Mean value	747.770
Standard Deviation	95.973

2. Yule's 'K' based on the joint occurrence of x_n, x_{n+1} :

Mean value	87.930
Standard Deviation	31.618

3. Yule's 'K' based on the joint occurrence of x_n, \dots, x_{n+2} :

Mean value	73.139
Standard Deviation	27.466

4.. Vowel-Consonant relationships are displayed in Table II.

TABLE X. ENGLISH (V-2) INDEPENDENT PROBABILITY OF X_n

X_n	Probability
A	.0659
E	.1000
I	.0585
O	.0638
U	.0234
B	.0114
C	.0258
D	.0344
F	.0183
G	.0179
H	.0414
J	.0014
K	.0059
L	.0367
M	.0210
N	.0603
P	.0173
Q	.0006
R	.0520
S	.0528
T	.0725
V	.0093
W	.0155
X	.0016
Y	.0157
Z	.0007
Space	.1760

TABLE XI. ENGLISH (V-2) JOINT PROBABILITY TABLE FOR X_n , X_{n+1}

		X_n												
		A	E	I	O	U	B	C	D	F	G	H	J	K
A	.0000	.0047	.0010	.0008	.0007	.0014	.0042	.0014	.0010	.0012	.0062	.0001	.0001	
E	.0000	.0033	.0020	.0002	.0008	.0030	.0030	.0037	.0012	.0026	.0176	.0003	.0019	
I	.0022	.0013	.0000	.0007	.0006	.0008	.0014	.0025	.0025	.0009	.0056	.0000	.0008	
O	.0000	.0005	.0040	.0016	.0000	.0014	.0055	.0025	.0037	.0010	.0042	.0004	.0001	
U	.0007	.0001	.0000	.0081	.0000	.0015	.0017	.0008	.0007	.0005	.0112	.0006	.0000	
B	.0013	.0002	.0007	.0005	.0004	.0001	.0001	.0000	.0000	.0000	.0011	.0000	.0000	
C	.0020	.0026	.0043	.0011	.0011	.0000	.0002	.0000	.0000	.0000	.0000	.0000	.0000	
D	.0025	.0086	.0023	.0011	.0005	.0000	.0000	.0003	.0000	.0000	.0000	.0000	.0000	
F	.0008	.0008	.0013	.0049	.0001	.0000	.0000	.0000	.0011	.0000	.0000	.0000	.0000	
G	.0012	.0006	.0025	.0005	.0010	.0000	.0000	.0001	.0000	.0002	.0000	.0000	.0000	
H	.0000	.0001	.0000	.0002	.0000	.0000	.0035	.0000	.0000	.0023	.0000	.0000	.0000	
J	.0001	.0000	.0000	.0001	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	
K	.0008	.0002	.0003	.0006	.0000	.0000	.0017	.0000	.0000	.0000	.0000	.0000	.0000	
L	.0069	.0035	.0033	.0034	.0020	.0013	.0009	.0003	.0005	.0004	.0001	.0000	.0002	
M	.0020	.0020	.0024	.0034	.0008	.0000	.0002	.0000	.0000	.0000	.0000	.0000	.0000	
N	.0130	.0097	.0154	.0110	.0032	.0000	.0001	.0001	.0000	.0005	.0002	.0000	.0004	
P	.0008	.0011	.0007	.0016	.0011	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	
Q	.0000	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	
R	.0077	.0122	.0022	.0082	.0037	.0005	.0015	.0014	.0013	.0013	.0007	.0000	.0000	
S	.0056	.0030	.0055	.0017	.0034	.0001	.0009	.0001	.0005	.0001	.0001	.0000	.0004	
T	.0087	.0029	.0060	.0025	.0030	.0001	.0020	.0000	.0012	.0012	.0015	.0000	.0000	
V	.0014	.0021	.0023	.0012	.0000	.0000	.0001	.0001	.0000	.0000	.0000	.0000	.0000	
W	.0006	.0009	.0000	.0024	.0000	.0000	.0001	.0001	.0000	.0000	.0000	.0000	.0000	
X	.0002	.0009	.0004	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	
Y	.0019	.0009	.0000	.0001	.0001	.0010	.0010	.0003	.0004	.0000	.0003	.0002	.0000	
Z	.0001	.0000	.0004	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	
Space	.0055	.0329	.0016	.0077	.0008	.0002	.0009	.0194	.0051	.0062	.0036	.0000	.0021	

TABLE XI - Continued

	X_n	X_{n+1}	L	M	N	P	Q	R	S	T	V	W	X	Y	Z	Space
A	.0034	.0022	.0019	.0000	.0040	.0023	.0025	.0005	.0030	.0001	.0001	.0001	.0001	.0195		
E	.0062	.0053	.0058	.0031	.0000	.0113	.0052	.0083	.0066	.0031	.0001	.0008	.0003	.0045		
I	.0045	.0027	.0026	.0008	.0000	.0043	.0031	.0071	.0017	.0022	.0001	.0002	.0001	.0100		
O	.0023	.0024	.0024	.0022	.0000	.0050	.0021	.0072	.0004	.0020	.0000	.0014	.0000	.0114		
U	.0006	.0004	.0006	.0006	.0007	.0017	.0012	.0000	.0000	.0000	.0000	.0000	.0000	.0020		
B	.0000	.0005	.0000	.0000	.0001	.0000	.0002	.0000	.0000	.0000	.0000	.0001	.0000	.0072		
C	.0000	.0001	.0019	.0000	.0000	.0009	.0008	.0002	.0000	.0001	.0000	.0000	.0000	.0105		
D	.0020	.0000	.0098	.0000	.0000	.0013	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0059		
F	.0004	.0000	.0003	.0000	.0000	.0001	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0084		
G	.0000	.0000	.0074	.0000	.0000	.0012	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0032		
H	.0000	.0000	.0001	.0004	.0000	.0001	.0021	.0194	.0000	.0023	.0001	.0000	.0000	.0107		
J	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0012		
K	.0002	.0000	.0005	.0000	.0000	.0007	.0003	.0000	.0000	.0000	.0000	.0000	.0000	.0007		
L	.0057	.0001	.0003	.0015	.0000	.0005	.0005	.0007	.0007	.0000	.0001	.0000	.0004	.0043		
M	.0001	.0005	.0003	.0000	.0000	.0008	.0008	.0007	.0002	.0000	.0000	.0000	.0000	.0077		
N	.0000	.0000	.0005	.0000	.0000	.0012	.0001	.0000	.0005	.0000	.0001	.0000	.0000	.0043		
P	.0003	.0016	.0001	.0010	.0000	.0003	.0011	.0000	.0000	.0000	.0003	.0000	.0000	.0074		
Q	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0003		
R	.0001	.0001	.0000	.0030	.0000	.0007	.0000	.0026	.0000	.0001	.0000	.0000	.0000	.0046		
S	.0010	.0005	.0025	.0004	.0000	.0039	.0023	.0019	.0000	.0003	.0000	.0004	.0000	.0132		
T	.0005	.0000	.0065	.0004	.0000	.0023	.0066	.0011	.0000	.0005	.0001	.0001	.0000	.0264		
V	.0002	.0000	.0006	.0000	.0000	.0003	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0011		
W	.0000	.0000	.0001	.0000	.0000	.0001	.0003	.0011	.0000	.0000	.0000	.0000	.0000	.0098		
X	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
Y	.0027	.0005	.0007	.0002	.0000	.0014	.0002	.0027	.0001	.0000	.0000	.0000	.0000	.0019		
Z	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000		
Space	.0064	.0026	.0153	.0016	.0000	.0108	.0230	.0161	.0000	.0019	.0004	.0118	.0000			

TABLE XII. ENGLISH (V-2) JOINT PROBABILITY TABLE FOR X_n , -, X_{n+2}

		X_n												
		A	E	I	O	U	B	C	D	F	G	H	J	K
A	.0011	.0048	.0029	.0022	.0016	.0004	.0023	.0029	.0009	.0013	.0009	.0000	.0000	.0004
E	.0084	.0079	.0081	.0075	.0021	.0012	.0020	.0021	.0015	.0012	.0008	.0000	.0000	.0003
I	.0045	.0054	.0033	.0026	.0015	.0007	.0018	.0017	.0010	.0010	.0009	.0000	.0000	.0002
O	.0015	.0037	.0007	.0009	.0005	.0004	.0011	.0015	.0015	.0011	.0009	.0000	.0000	.0005
U	.0005	.0014	.0012	.0007	.0002	.0006	.0015	.0004	.0006	.0001	.0015	.0000	.0000	.0001
B	.0004	.0016	.0002	.0006	.0003	.0001	.0002	.0013	.0001	.0003	.0002	.0001	.0001	.0001
C	.0017	.0036	.0008	.0010	.0006	.0007	.0001	.0018	.0011	.0004	.0007	.0001	.0001	.0001
D	.0085	.0028	.0013	.0022	.0028	.0002	.0004	.0016	.0004	.0004	.0011	.0001	.0001	.0006
F	.0007	.0020	.0003	.0009	.0002	.0002	.0002	.0011	.0005	.0003	.0002	.0000	.0000	.0002
G	.0012	.0021	.0060	.0021	.0005	.0003	.0002	.0003	.0002	.0001	.0004	.0000	.0000	.0000
H	.0012	.0028	.0035	.0014	.0013	.0000	.0001	.0014	.0003	.0004	.0003	.0001	.0001	.0001
J	.0000	.0002	.0000	.0001	.0000	.0000	.0000	.0002	.0000	.0000	.0000	.0000	.0000	.0000
K	.0012	.0006	.0009	.0014	.0002	.0000	.0000	.0001	.0000	.0001	.0000	.0000	.0000	.0000
L	.0035	.0024	.0023	.0043	.0013	.0007	.0018	.0017	.0017	.0004	.0017	.0000	.0000	.0002
M	.0009	.0022	.0006	.0015	.0001	.0000	.0016	.0011	.0004	.0003	.0017	.0001	.0001	.0001
N	.0015	.0029	.0047	.0026	.0006	.0007	.0041	.0021	.0006	.0010	.0033	.0001	.0007	.0007
P	.0013	.0032	.0007	.0015	.0005	.0000	.0004	.0008	.0003	.0003	.0007	.0000	.0001	.0001
Q	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000
R	.0012	.0038	.0003	.0029	.0004	.0008	.0012	.0016	.0038	.0011	.0031	.0002	.0003	.0003
S	.0059	.0072	.0024	.0051	.0010	.0004	.0014	.0021	.0007	.0012	.0038	.0004	.0004	.0003
T	.0043	.0112	.0035	.0051	.0024	.0015	.0007	.0039	.0016	.0034	.0031	.0000	.0004	.0004
V	.0002	.0009	.0005	.0003	.0001	.0001	.0001	.0004	.0008	.0004	.0009	.0000	.0000	.0000
W	.0004	.0024	.0003	.0004	.0001	.0001	.0011	.0003	.0003	.0003	.0009	.0000	.0002	.0002
X	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0001	.0001	.0001	.0000	.0000	.0000	.0000
Y	.0014	.0012	.0010	.0005	.0001	.0003	.0001	.0009	.0003	.0002	.0007	.0000	.0002	.0002
Z	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Space	.0161	.0239	.0129	.0164	.0049	.0020	.0045	.0024	.0007	.0026	.0136	.0001	.0011	.0011

TABLE XII - Continued

	L	M	N	P	Q	R	S	T	V	W	X	Y	Z	Space
A	.0027	.0011	.0031	.0010	.0003	.0034	.0064	.0050	.0001	.0006	.0001	.0020	.0000	.0186
E	.0021	.0010	.0045	.0022	.0002	.0045	.0034	.0156	.0001	.0010	.003	.0008	.0000	.0211
I	.0013	.0007	.0031	.0015	.0002	.0025	.0041	.0036	.0001	.0010	.0001	.0009	.0000	.0151
O	.0016	.0005	.0023	.0027	.0000	.0018	.0039	.0065	.0001	.0007	.0001	.0008	.0000	.0288
U	.0004	.0003	.0007	.0002	.0000	.0011	.0013	.0008	.0000	.0006	.0000	.0013	.0000	.0081
B	.0005	.0001	.0006	.0003	.0000	.0011	.0012	.0008	.0000	.0001	.0000	.0005	.0000	.0008
C	.0014	.0007	.0012	.0006	.0000	.0029	.0019	.0021	.0003	.0002	.0001	.0007	.0000	.0012
D	.0012	.0010	.0012	.0005	.0000	.0027	.0016	.0022	.0005	.0003	.0001	.0006	.0001	.0005
F	.0008	.0002	.0007	.0001	.0000	.0010	.0015	.0008	.0000	.0002	.0001	.0009	.0000	.0056
G	.0009	.0004	.0003	.0001	.0000	.0010	.0005	.0006	.0000	.0001	.0000	.0002	.0000	.0003
H	.0005	.0002	.0015	.0002	.0000	.0015	.0014	.0013	.0000	.0002	.0001	.0007	.0000	.0211
J	.0001	.0001	.0001	.0000	.0000	.0002	.0002	.0001	.0000	.0000	.0000	.0001	.0000	.0000
K	.0003	.0005	.0001	.0001	.0000	.0001	.0000	.0002	.0000	.0000	.0000	.0000	.0000	.0001
L	.0002	.0014	.0017	.0011	.0000	.0017	.0018	.0021	.0007	.0010	.0010	.0003	.0000	.0038
M	.0007	.0002	.0010	.0001	.0000	.0025	.0019	.0018	.0001	.0002	.0001	.0005	.0000	.0015
N	.0019	.0034	.0019	.0011	.0000	.0022	.0025	.0028	.0013	.0012	.0001	.0008	.0002	.0162
P	.0006	.0001	.0007	.0002	.0000	.0015	.0012	.0008	.0000	.0001	.0000	.0004	.0000	.0018
Q	.0000	.0000	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
R	.0021	.0021	.0030	.0027	.0000	.0005	.0024	.0048	.0023	.0018	.0001	.0005	.0001	.0090
S	.0025	.0019	.0029	.0009	.0000	.0036	.0021	.0022	.0007	.0016	.0001	.0012	.0000	.0031
T	.0025	.0019	.0065	.0008	.0000	.0035	.0038	.0034	.0004	.0016	.0001	.0017	.0001	.0054
V	.0004	.0002	.0004	.0000	.0000	.0009	.0008	.0005	.0001	.0002	.0000	.0001	.0000	.0011
W	.0010	.0002	.0016	.0002	.0000	.0011	.0017	.0010	.0000	.0001	.0000	.0006	.0000	.0016
X	.0000	.0000	.0002	.0000	.0000	.0004	.0004	.0002	.0000	.0000	.0000	.0000	.0000	.0007
Y	.0010	.0003	.0013	.0004	.0000	.0013	.0006	.0011	.0000	.0007	.0002	.0001	.0000	.0021
Z	.0010	.0000	.0002	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Space	.0100	.0027	.0197	.0066	.0000	.0094	.0061	.0123	.0027	.0024	.0001	.0003	.0001	.0084

APPENDIX D. SPANISH POPULATION VALUES (V-2)

1. Yule's 'K' based on the independent occurrence of each character

x_n :

Mean value 806.481

Standard Deviation 77.310

2. Yule's 'K' based on the joint occurrence of x_n, x_{n+1} :

Mean value 108.784

Standard Deviation 30.665

3. Yule's 'K' based on the joint occurrence of x_n, \dots, x_{n+2} :

Mean value 83.553

Standard Deviation 22.886

4. Vowel-Consonant relationships are displayed in Table VI.

TABLE XIII. SPANISH (V-2) INDEPENDENT PROBABILITY OF X_n

X_n	Probability
A	.1053
E	.1088
I	.0573
O	.0771
U	.0339
B	.0114
C	.0372
D	.0390
F	.0060
G	.0092
H	.0072
J	.0034
K	.0005
L	.0471
M	.0236
N	.0620
P	.0205
Q	.0073
R	.0514
S	.0625
T	.0385
V	.0095
W	.0000
X	.0011
Y	.0074
Z	.0031
Space	.1699

TABLE XIV. SPANISH (V-2) JOINT PROBABILITY TABLE FOR X_n , X_{n+1}

		X_n													
		I	O	U	B	C	D	F	G	H	J	K			
A	.0000	.0011	.0076	.0001	.0021	.0036	.0070	.0050	.0006	.0013	.0033	.0008	.0008	.0000	
E	.0002	.0002	.0071	.0000	.0096	.0008	.0033	.0183	.0010	.0011	.0008	.0009	.0009	.0001	
I	.0006	.0006	.0000	.0002	.0019	.0022	.0087	.0044	.0017	.0011	.0009	.0001	.0002		
O	.0002	.0007	.0068	.0002	.0002	.0008	.0088	.0079	.0010	.0016	.0018	.0012	.0000		
U	.0008	.0004	.0001	.0000	.0000	.0005	.0051	.0011	.0009	.0019	.0004	.0004	.0000		
B	.0043	.0004	.0011	.0013	.0005	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
C	.0054	.0041	.0046	.0022	.0014	.0000	.0007	.0000	.0000	.0000	.0000	.0000	.0000		
D	.0067	.0025	.0035	.0017	.0010	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
F	.0003	.0009	.0005	.0005	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
G	.0005	.0019	.0014	.0007	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
H	.0002	.0000	.0000	.0000	.0000	.0000	.0023	.0000	.0000	.0000	.0000	.0000	.0000		
J	.0006	.0009	.0003	.0003	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
K	.0000	.0001	.0000	.0001	.0000	.0000	.0001	.0001	.0000	.0000	.0000	.0000	.0000		
L	.0077	.0092	.0032	.0024	.0013	.0012	.0006	.0000	.0003	.0002	.0000	.0000	.0001		
M	.0036	.0023	.0027	.0035	.0008	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
N	.0116	.0200	.0063	.0108	.0073	.0000	.0001	.0001	.0000	.0002	.0000	.0000	.0000		
P	.0010	.0007	.0003	.0009	.0005	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
Q	.0004	.0002	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
R	.0114	.0107	.0018	.0067	.0020	.0017	.0006	.0007	.0005	.0015	.0000	.0000	.0000		
S	.0104	.0155	.0033	.0152	.0021	.0003	.0000	.0000	.0000	.0000	.0000	.0000	.0001		
T	.0025	.0025	.0023	.0014	.0007	.0001	.0019	.0000	.0000	.0000	.0000	.0000	.0000		
V	.0006	.0010	.0016	.0015	.0002	.0000	.0001	.0000	.0001	.0000	.0000	.0000	.0000		
W	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
X	.0000	.0010	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
Y	.0006	.0002	.0000	.0003	.0004	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
Z	.0006	.0010	.0005	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000		
Space	.0352	.0311	.0022	.0268	.0013	.0000	.0014	.0000	.0000	.0000	.0000	.0000	.0001		

TABLE XIV - Continued

TABLE XV. SPANISH (V-2) JOINT PROBABILITY TABLE FOR X_n , -, X_{n+2}

	X_n												
	A	E	I	O	U	B	C	D	F	G	H	J	K
A	.0123	.0095	.0086	.0074	.0054	.0016	.0047	.0015	.0004	.0012	.0000	.0001	.0000
E	.0102	.0090	.0031	.0078	.0016	.0024	.0026	.0013	.0010	.0001	.0001	.0001	.0000
I	.0092	.0074	.0049	.0051	.0019	.0008	.0022	.0003	.0002	.0007	.0000	.0000	.0000
O	.0088	.0059	.0052	.0045	.0027	.0004	.0043	.0008	.0001	.0008	.0000	.0000	.0001
U	.0032	.0029	.0014	.0018	.0003	.0001	.0007	.0000	.0001	.0001	.0000	.0000	.0000
B	.0011	.0007	.0002	.0008	.0001	.0001	.0005	.0004	.0001	.0000	.0013	.0001	.0000
C	.0042	.0051	.0011	.0021	.0011	.0002	.0004	.0015	.0012	.0002	.0010	.0001	.0000
D	.0056	.0037	.0009	.0051	.0012	.0002	.0012	.0020	.0000	.0003	.0001	.0001	.0000
F	.0008	.0007	.0003	.0005	.0001	.0000	.0001	.0004	.0000	.0000	.0000	.0001	.0000
G	.0016	.0007	.0004	.0008	.0005	.0000	.0001	.0000	.0001	.0000	.0001	.0000	.0000
H	.0011	.0014	.0001	.0014	.0006	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
J	.0002	.0002	.0002	.0000	.0003	.0000	.0003	.0000	.0000	.0000	.0001	.0000	.0000
K	.0001	.0001	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0000
L	.0049	.0059	.0025	.0022	.0012	.0004	.0019	.0021	.0002	.0003	.0000	.0001	.0000
M	.0029	.0024	.0016	.0018	.0003	.0001	.0028	.0007	.0001	.0007	.0002	.0000	.0000
N	.0015	.0020	.0082	.0010	.0016	.0008	.0056	.0018	.0008	.0013	.0005	.0003	.0000
P	.0030	.0045	.0009	.0025	.0004	.0000	.0007	.0002	.0000	.0000	.0000	.0000	.0000
Q	.0015	.0010	.0002	.0015	.0002	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
R	.0029	.0028	.0019	.0020	.0016	.0010	.0027	.0022	.0009	.0011	.0006	.0004	.0000
S	.0033	.0039	.0026	.0029	.0015	.0004	.0025	.0044	.0003	.0005	.0008	.0005	.0000
T	.0061	.0133	.0028	.0029	.0019	.0004	.0004	.0006	.0002	.0003	.0000	.0001	.0000
V	.0012	.0013	.0004	.0007	.0005	.0000	.0001	.0004	.0000	.0000	.0004	.0004	.0000
W	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
X	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
Y	.0015	.0008	.0001	.0010	.0000	.0000	.0001	.0001	.0000	.0000	.0002	.0000	.0001
Z	.0004	.0003	.0004	.0000	.0001	.0000	.0001	.0000	.0000	.0000	.0002	.0000	.0000
Space	.0176	.0235	.0095	.0210	.0094	.019	.0036	.0181	.0002	.0013	.0011	.0008	.0001

TABLE XV - Continued

	X_n						X_{n+2}							
	L	M	N	P	Q	R	S	T	V	W	X	Y	Z	Space
A	.0032	.0008	.0061	.0011	.0000	.0055	.0064	.0026	.0004	.0000	.0005	.0001	.0263	
E	.0023	.0012	.0100	.0028	.0060	.0048	.0054	.0018	.0008	.0000	.0001	.0001	.0326	
I	.0010	.0008	.0040	.0012	.0012	.0024	.0018	.0005	.0002	.0000	.0000	.0000	.0116	
O	.0017	.0007	.0060	.0016	.0000	.0039	.0032	.0017	.0004	.0000	.0002	.0002	.0239	
U	.0005	.0001	.0024	.0002	.0000	.0009	.0018	.0003	.0000	.0000	.0002	.0000	.0171	
B	.0011	.0001	.0004	.0002	.0000	.0012	.0009	.0008	.0001	.0000	.0000	.0001	.0009	
C	.0029	.0008	.0032	.0012	.0000	.0028	.0029	.0021	.0007	.0000	.0000	.0005	.0018	
D	.0020	.0012	.0026	.0009	.0000	.0019	.0054	.0025	.0006	.0000	.0000	.0006	.0010	
F	.0045	.0000	.0004	.0000	.0000	.0006	.0006	.0003	.0000	.0000	.0000	.0001	.0006	
G	.0011	.0002	.0004	.0000	.0000	.0010	.0010	.0005	.0001	.0000	.0000	.0001	.0004	
H	.0003	.0000	.0005	.0000	.0000	.0002	.0008	.0000	.0000	.0000	.0000	.0002	.0007	
J	.0003	.0003	.0002	.0000	.0000	.0003	.0004	.0001	.0000	.0000	.0000	.0000	.0003	
K	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	
L	.0005	.0022	.0032	.0009	.0000	.0017	.0035	.0015	.0006	.0000	.0002	.0007	.0100	
M	.0016	.0002	.0014	.0000	.0000	.0016	.0016	.0015	.0002	.0000	.0002	.0004	.0013	
N	.0020	.0053	.0014	.0011	.0000	.0034	.0033	.0033	.0010	.0000	.0001	.0003	.0152	
P	.0014	.0000	.0010	.0000	.0000	.0014	.0029	.0001	.0000	.0000	.0000	.0003	.0008	
Q	.0003	.0001	.0005	.0000	.0000	.0004	.0010	.0000	.0001	.0000	.0001	.0004		
R	.0018	.0023	.0033	.0064	.0000	.0009	.0032	.0036	.0013	.0000	.0002	.0004	.0076	
S	.0083	.0027	.0052	.0013	.0000	.0056	.0042	.0036	.0011	.0000	.0001	.0008	.0056	
T	.0014	.0010	.0016	.0004	.0000	.0015	.0016	.0005	.0001	.0000	.0002	.0000	.0011	
V	.0007	.0002	.0008	.0000	.0000	.0005	.0011	.0008	.0001	.0000	.0000	.0000	.0003	
W	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	
X	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0000	.0000	.0000	.0000	.0009	
Y	.0003	.0005	.0004	.0001	.0000	.0004	.0015	.0001	.0001	.0000	.0001	.0002	.0001	
Z	.0003	.0000	.0001	.0001	.0000	.0004	.0000	.0003	.0001	.0000	.0000	.0000	.0001	
Space	.0117	.0031	.0071	.0010	.0000	.0082	.0076	.0103	.0014	.0000	.0010	.0010	.0095	

APPENDIX E. TABLE XVI - Significance Levels for Kolmogorov-Smirnov Test

α	z	α	z								
0.28	.999999	0.59	.877240	0.90	.392730	1.21	.106970	1.52	.019690		
0.29	.999996	0.60	.864282	0.91	.379072	1.22	.101896	1.53	.018524		
0.30	.999991	0.61	.850771	0.92	.365714	1.23	.097028	1.54	.017422		
0.31	.999979	0.62	.836775	0.93	.352662	1.24	.092352	1.55	.016378		
0.32	.999954	0.63	.822247	0.94	.339918	1.25	.087868	1.56	.015390		
0.33	.999909	0.64	.807333	0.95	.327484	1.26	.083568	1.57	.014456		
0.34	.999829	0.65	.792013	0.96	.315364	1.27	.079444	1.58	.013574		
0.35	.999697	0.66	.776363	0.97	.303556	1.28	.075495	1.59	.012740		
0.36	.999489	0.67	.760418	0.98	.292060	1.29	.071712	1.60	.011952		
0.37	.999174	0.68	.744220	0.99	.280874	1.30	.068092	1.61	.011209		
0.38	.998715	0.69	.727811	1.00	.270000	1.31	.064630	1.62	.010508		
0.39	.998071	0.70	.711235	1.01	.259434	1.32	.061318	1.63	.009846		
0.40	.997192	0.71	.694529	1.02	.249174	1.33	.058152	1.64	.009223		
0.41	.996028	0.72	.677735	1.03	.239220	1.34	.055128	1.65	.008636		
0.42	.994524	0.73	.660887	1.04	.229566	1.35	.052244	1.66	.008083		
0.43	.992623	0.74	.644019	1.05	.220206	1.36	.049488	1.67	.007562		
0.44	.990270	0.75	.627167	1.06	.211140	1.37	.046858	1.68	.007072		
0.45	.987410	0.76	.610360	1.07	.202364	1.38	.044350	1.69	.006611		
0.46	.983995	0.77	.593628	1.08	.193872	1.39	.041960	1.70	.006177		
0.47	.979978	0.78	.576998	1.09	.185658	1.40	.039682	1.71	.005770		
0.48	.975318	0.79	.560495	1.10	.177718	1.41	.037514	1.72	.005388		
0.49	.969983	0.80	.544143	1.11	.170050	1.42	.035448	1.73	.005028		
0.50	.963645	0.81	.527959	1.12	.162644	1.43	.033484	1.74	.004691		
0.51	.957186	0.82	.511970	1.13	.155498	1.44	.031618	1.75	.004375		
0.52	.949694	0.83	.496192	1.14	.148606	1.45	.029842	1.76	.004078		
0.53	.941466	0.84	.480634	1.15	.141962	1.46	.028154	1.77	.003800		
0.54	.932503	0.85	.465318	1.16	.135558	1.47	.026552	1.78	.003540		
0.55	.922817	0.86	.450256	1.17	.129388	1.48	.025030	1.79	.003296		
0.56	.912423	0.87	.435454	1.18	.123452	1.49	.023588	1.80	.003068		
0.57	.901344	0.88	.420930	1.19	.117742	1.50	.022218	1.81	.002845		
0.58	.889605	0.89	.406684	1.20	.112250	1.51	.020920	1.82	.002645		

APPENDIX F. FINAL TEST SOURCES

From Introduction to Literature: Stories [Ref. 10]:

- Anderson, Sherwood, Death in the Woods.
Benét, Stephen Vincent, By The Waters of Babylon.
Crane, Stephen, The Open Boat.
Curley, Daniel, A Story of Love, Etc.
Edmonds, Walter D., Death of Red Peril.
Fitzgerald, F. Scott, Absolution.
Forster, E. M., The Other Side of the Hedge.
Galsworthy, John, Quality.
Hardy, Thomas, The Three Strangers.
Porter, Katherine Anne, The Grave.
Schwartz, Delmore, In Dreams Begin Responsibilities.
Steele, Wilbur Daniel, Footfalls.
Steinbeck, John, from The Red Pony: The Gift.
Stevenson, Robert Louis, Markheim.
White, E. B., The Dove.

From Five Centuries of Spanish Literature [Ref. 11]:

- Anonymous, Lazarillo de Tormes.
Cervantes Saavedra, Miguel De, El ingenioso hidalgo, don Quijote de la Mancha.
Manuel, Don Juan, El Conde Lucanor.

From Antología Venezolana [Ref. 12]:

- Alfonzo, Alfredo Armas, La hora y punto.
Arráiz, Antonio, El mar es como un potro vigoroso.

Coll, Pedro Emilio, El diente roto.
Croce, Arturo, Angustia apagada.
Diaz Rodriguez, Manuel, Mi Alma era una mina abandonada.
Diaz Sanchez, Ramon, Poderes de la mas alta valia.
Diaz Solis, Gustavo, La efigie.
Fombona, Rufino Blanco, El catire.
Gallegos, Romulo, Canaima.
Garmendia, Julio, Eladia.
Guaramato, Oscar, Vecindad..
La Parra, Teresa De, Vicente Cochicho.
Marquez Salas, Antonio, Como Dios!.
Meneses, Guillermo, El companero Juan Ruiz.
Mujica, Hector, Las tres ventanas.
Núñez, Enrique Bernardo, Vocchi.
Palacios, Lucila, Cap. XX (Cubil).
Picon Salas, Mariano, Josefita.
Pocaterra, Jose Rafael, Patria, la mestiza.
Pardo, Isaac J., El domador de potros.
Pardon, Julian, II(Primavera nocturna).
Rivas Mijares, Humberto, El murado.
Rosales, Julio, El mejor rabula.
Ruiz, Jose Fabbiani, A orillas del viejo rio.
Silva, Miguel Otero, Fiebre.
Stolk, Gloria, Cap. I (Amargo fondo).
Urbaneja Achelpohl, Luis M., Ovejon...!.
Ustar Pietri, Arturo, X (Las lanzas coloradas).

From An Anthology of Spanish American Literature [Ref. 13]:

Altamirano, Ignacio Manuel, El dia de muertos.

Bastamente Carlos Inga, Calixto (Concolorcorvo), El lazarillo de ciegos caminantes.

Bello, Andres, Silva a la Argicultura de la Zona Torrida.

Blanco-Fombona, Rufino, El conquistador español del siglo XVI.

Bolivar, Simon, Carta a un caballero que tomaba gran interes en la causa republicana en la america del sur.

Chocano, Jose Santos, La Epopeya Del Pacifico.

Cortes, Hernan, The Death of Moctezuma and the Tragic Retreat of the Spaniards from the Aztec Capital the Night of June 30, 1520.

Dario, Ruben, Azul (El rey burgues).

Diaz del Castillo, Bernal, The Imprisonment of Moctezuma.

Fernandez de Lizardi, Jose Joaquin, El periquillo sarniento.

Garcia Calderon, Francisco, La creacion de un continente.

Garcilaso de la Vega, El Inca, Comentarios reales de los incas.

Gonzalez Prada, Manuel, Discurso en el politeama.

Gutierrez Najera, Manuel, La Novela Del Tranvia.

Hernandez, Jose, Martin Fierro.

Juana Ines de la Cruz, Sor, Respuesta a Sor Filotea de la Cruz.

Latorre, Mariano, Chilenos del mar (El piloto Oyarzo).

Lillo, Baldomero, El chiflon del diablo.

Lopez Albujar, Enrique, Cuentos andinos (Como habla la coca).

Marti, Jose, Nuestra America.

Mitre, Bartolome, La historia de San Martin.

Montalvo, Juan, Washington y Bolivar.

Nervo, Amado, Una esperanza.

Palma, Ricardo, Las orejas del Alcalde.

Quiroga, Horacio, El desierto.

Reyes, Alfonso, Vision de Anahuac.
Rodo, Jose Enrique, Ariel.
Rojas, Manuel, Hombres del sur (El anchorro).
Rojas, Ricardo, La argentinidad.
Sarmiento, Domingo Faustino, Facundo.
Siguenza y Gongora, Carlos de, Infortunios de Alonso Rameriz.
Ugarte, Manuel, El destino de un continente (La neuva Roma).
Vasconcelos, Jose, La raza cosmica.
Viana, Javier de, Leña seca (El domador).

From The Creative Reader [Ref. 14]:

Algren, Nelson, The Moon of the Arfy Darfy.
Amster, L. J., Center of Gravity.
Auchincloss, Louis, Power in Trust.
Berriault, Gina, The Birthday Party.
Boyle, Kay, The Ballet of Central Park.
Buechler, James, John Sobieski Runs.
Calisher, Hortense, Little Did I Know.
Connell, Evan S., Jr., The Suicide.
Cozzens, James Gould, One Hundred Ladies.
Elkin, Stanley, Perlmutter at the East Pole.
Gary, Romain, A Humanist.
Gold, Hebert, Dance of the Divorced.
Goyen, William, Figure over the Town.
Greene, Graham, The Root of All Evil.
Malamud, Bernard, The Refugee.
McCullers, Carson, Sucker.
Miller, Warren, Chaos, Disorder and the Late Show.

Moravia, Alberto, A Tough Nut.
O'Connor, Frank, A Life of Your Own.
O'Faolain, Sean, One Man, One Boat, One Girl.
Pynchon, Thomas, The Secret Integration.
Saroyan, William, Boys and Girls Together.
Seager, Allan, No More Roses.
Shaw, Irwin, The Inhabitants of Venus.
Singer, Isaac Bashevis, Esther Kreindel The Second.
Swados, Harvey, A Story For Teddy.
Updike, John, The Lucid Eye in Silver Town.
White, Robin, Walker's Peak.
Williams, Thomas, The Old Dancers.
Wood, Malcolm, The Appraiser.

From Best Modern Short Stories [Ref. 15]:

Aiken, Conrad, Mr. Arcularis.
Audry, Colette, The Gloves.
Bierce, Ambrose, An Occurrence at Owl Creek Bridge.
Bowen, Elizabeth, The Demon Lover.
Conrad, Hoseph, Amy Foster.
Faulkner, William, A Rose For Emily.
Granberry, Edwin, A Trip to Czardis.
Hawthorne, Nathaniel, My Kinsman, Major Molineux.
Hemingway, Ernest, A Clean, Well-Lighted Place.
James, Henry, Paste.
Joyce, James, A Little Cloud.
Lawrence, D. H., The Horse Dealer's Daughter.
Lowry, Malcolm, Strange Comfort Afforded by the Profession.

Mansfield, Katherine, The Fly.
Maupassant, Guy de, Two Friends.
Poe, Edgar Allan, Ligeia.
Taylor, Elizabeth, The First Death of Her Life.
Thurber, James, The Catbird Seat.
Welty, A Still Moment.
West, Jessamyn, Sixteen.

APPENDIX G. POPULATION ANALYSIS PROGRAM (V-1)

```

IDENTIFICATION DIVISION.
  V1-LANGUAGE-ANALYSIS.
  PPROGRAM-ID. RAU-USN.
  AUTHOR. LCDR. MORTON D. RAU.
  REMARKS. THIS PROGRAM IS PART OF MY MASTER OF SCIENCE IN
  MANAGEMENT THESIS AND IS DESIGNED TO CALCULATE VARIOUS
  STATISTICAL CHARACTERISTICS FOR LETTERS APPEARING IN
  LANGUAGES WRITTEN IN THE MODERN LATIN ALPHABET.

ENVIRONMENT DIVISION.
  LANGUAGE-COMPUTER. IBM-360-67.
  SOURCE-COMPUTER. IBM-360-67.

CONFIGURATION SECTION.
  SUBJECT-COMPUTER. IBM-360-67.
  SPECIAL-NAMES.
  CO1 IS TOP-OFF-PAGE.
  INPUT-OUTPUT SECTION.
  FILE-CONTROL.
    SELECT POPULATION ASSIGN TO UR-S-IN1.
    SELECT LANG-ANAL ASSIGN TO UR-S-OUT1.
    SELECT CARD-FILE ASSIGN TO UR-S-OUT2.

DATA DIVISION.
  FILE SECTION.
    FD PJPUEL RECORDS ARE OMITTED BLOCK CONTAINS 5 RECORDS
    DATA RECORD IS PJP-CARD, LANG-CARD.
    01 PJP-CARD.
      05 FILLER PIC XX.
    05 SAMPLE-FLD.
      10 CB-PIC X(7) OCCURS 71 TIMES.
    05 FILLER PIC X(7).

    01 LANG-CARD-DTYPE PIC A.
      05 LANG-KEY PIC X(10).
      05 NEW-LANG VALUE 'L'.
      05 FILLER PIC X(30).
      05 LANGUAGE PIC X(30).
      05 FILLER PIC X(38).
      05 LANG-KEY PIC X(10).

    FD LANG-ANAL RECORDS ARE OMITTED BLOCK CONTAINS 5 RECORDS
    DATA RECORD IS PRINT-LINE.
    01 PRINT-LINE PIC X(133).

    01 CARD-FILE RECORDS ARE OMITTED RECORD CONTAINS 80 CHARACTERS
    DATA RECORD IS CARD-OUT.
    01 CARD-OUT PIC X(80).

WORKING-STORAGE SECTION.

```

10 11 12 13 14 15 16 17 18 19 20

004650
 004700
 00048300
 00048500
 00049000
 00049500
 00050000
 00051000
 00051500
 00052000
 00052500
 00053000
 00053500
 00054000
 00054500
 00055000
 00055500
 00056000
 00056500
 00057000
 00057500
 00058000
 00058500
 00059000
 00059500
 00060000
 00060500
 00061000
 00061500
 00062000
 00062500
 00063000
 00063500
 00064000
 00064500
 00065000
 00065500
 00066000
 00066500
 00067000
 00067500
 00068000
 00068500
 00069000
 00069500
 00070000

05 FILLER PIC X VALUE 'Y'.
 05 FILLER PIC X VALUE 'Z'.
 05 FILLER PIC X VALUE '6'.
 05 FILLER PIC X VALUE '7'.
 01 05 LTR-FREQ-TABLE OCCURS 33 TIMES.
 01 05 IND-FREQ-TABLE OCCURS 33 TIMES.
 01 05 IND-PROB-TABLE OCCURS 33 TIMES.
 01 05 DEPENDENCY-FREQ-TABLE OCCURS 33 TIMES.
 01 05 DEPENDENCY-FREQ-TABLE OCCURS 33 TIMES.
 C1 05 DEPENDENCY-PROB-TABLE OCCURS 33 TIMES.
 05 GIVEN-LTR OCCURS 33 TIMES.
 10 FOL-PRCB PIC 9V9(5), COMP, SYNC, OCCURS 66 TIMES.
 05 OUTP-TONE-A FILLER PIC X(43) VALUE SPACES.
 05 FILLER PIC X(48) VALUE SPACES.
 - 05 CF EACH LETTER FILLER PIC X(42) VALUE SPACES.
 05 LANG-LINE FILLER PIC X(54) VALUE SPACES.
 05 FILLER PIC X(11) VALUE LANGUAGE: .
 05 LANG-DES PIC X(30) VALUE SPACES.
 05 FILLER PIC X(38) VALUE SPACES.
 05 OUTP-CHN FILLER PIC X(37) VALUE SPACES.
 05 FILLER PIC X(25) VALUE LETTER: .
 05 FILLER PIC X(46) VALUE FREQUENCY: .
 05 OUTP-TONE-B FILLER PIC X(43) VALUE SPACES.
 05 FILLER PIC X(48) VALUE ALL: .
 05 FILLER PIC X(42) VALUE SPACES.
 05 LANG-FUNLINES FILLER PIC X(65) VALUE SPACES.
 05 FILLER PIC X(30) VALUE ALL: .
 05 FILLER PIC X(33) VALUE SPACES.
 05 OUTP-TONE-C FILLER PIC X(37) VALUE SPACES.
 05 FILLER PIC X(61) VALUE ALL: .
 05 FILLER PIC X(35) VALUE SPACES.
 C1 05 OUTP-TONE-D FILLER PIC X(40) VALUE SPACES.
 05 LTR-ONE FILLER PIC X(22) VALUE SPACES.
 05 FREQ-ONE FILLER PIC X(22) VALUE ZZZ, ZZ9.

```

007050 05 FILLER PIC X(19) VALUE SPACES.
007100 05 PROB-ONE PIC X(9)9999.
007150 05 FILLER PIC X(37) VALUE SPACES.
007200 01 05 OUTPT-ONE-F FILLER PIC X(38) VALUE SPACES.
007250 05 OUTPT-ONE-F FILLER PIC X(22) VALUE TOTAL.
007300 05 ONE-F-1 FILLER PIC ZZZZZZ.
007350 05 FILLER PIC X(63) VALUE SPACES.
007400 01 05 OUTPT-TWOC-A FILLER PIC X(46) VALUE SPACES.
007450 05 FILLER PIC X(87) VALUE SPACES.
007500 01 05 OUTPT-TWOC-B FILLER PIC X(11) VALUE SPACES.
007550 01 05 FIRST-FILLER PIC X(46) VALUE SPACES.
007600 01 05 FILLER PIC X(40) VALUE ALL.
007650 01 05 FIRST-OUTPT FILLER PIC X(47) VALUE SPACES.
007700 01 05 FILLER PIC X(102) VALUE SPACES.
007750 01 05 FIRST-LTR-TITLE FILLER PIC X(14) VALUE SPACES.
007800 01 05 FIRST-LTR-PIC X(9) OCCURS 13 TIMES.
007850 01 05 FILLER PIC X(1) VALUE SPACES.
007900 01 05 DEPN-FREQLINE FILLER PIC X(5) VALUE SPACES.
007950 01 05 DEPN-FREQLINE FILLER PIC X(1) XXX VALUE '(F)'.
008000 01 05 DEPN-FREQLINE FILLER PIC X(7) PIC X(BZZZZZ) OCCURS 13 TIMES.
008050 01 05 DEPN-FREQLINE FILLER PIC X(7) VALUE SPACES.
008100 01 05 DEPN-FREQLINE FILLER PIC X(5) VALUE SPACES.
008150 01 05 DEPN-FREQLINE FILLER PIC X(1) XXX VALUE '(P)'.
008200 01 05 DEPN-FREQLINE FILLER PIC X(7) PIC X(BB9999) OCCURS 13 TIMES.
008250 01 05 OUTPT-THREE-E FILLER PIC X(46) VALUE SPACES.
008300 01 05 OUTPT-THREE-E FILLER PIC X(87) VALUE SPACES.
008350 01 05 OUTPT-THREE-E FILLER PIC X(1) VALUE SPACES.
008400 01 05 OUTPT-THREE-E FILLER PIC X(47) VALUE SPACES.
008450 01 05 OUTPT-THREE-E FILLER PIC X(33) VALUE SPACES.
008500 01 05 OUTPT-THREE-E FILLER PIC X(9)9999. VALUE '(VOWEL/VOWEL)'.
008550 01 05 OUTPT-THREE-E FILLER PIC X(46) VALUE SPACES.
008600 01 05 OUTPT-THREE-E FILLER PIC X(47) VALUE SPACES.
008650 01 05 OUTPT-THREE-E FILLER PIC X(33) VALUE SPACES.

```

```

009450 01 P-CONS-VCH PIC X(46) VALUE 9.99999 SPACES.
009550 01 05 OUTPT-FOUR-CIC X(47) VALUE SPACES *P(VOWEL/CONSONANT)*.
009650 01 05 FILLER PIC X(33) VALUE SPACES *P(VOWEL/CONSONANT)*.
009750 01 05 P-VOW-CONNS PIC 9.99999 SPACES.
009850 01 05 OUTPT-FILLER-DIC X(46) VALUE SPACES.
009950 01 05 FILLER-PIC X(47) VALUE SPACES *P(CONS/CONSONANT/CONSONANT)*.
009960 01 05 FILCNS-CONNS PIC 9.99999 SPACES.
009970 01 05 OUTPT-FILLER-EIC X(33) VALUE SPACES *P(CONS/CONSONANT/CONSONANT)*.
009980 01 05 FILLER-PIC X(46) VALUE SPACES.
009990 01 05 OUTPT-FILLER-FIC X(11) VALUE SPACES *YULE-S CHARACTERISTIC, K, FOR: *.
010000 C1 05 FILLEF-PIC X(32) VALUE SPACES *YULE-S CHARACTERISTIC, K, FOR: *.
010010 01 05 OUTPT-FOUR-LANG PIC X(30).
010020 01 05 OUTPT-FOUR-LANG PIC X(12) 9.999 SPACES.
010030 01 05 OUTPT-FILLER-K PIC X(40) VALUE SPACES.
010040 01 05 OUTPT-FILLER-FIC X(54) VALUE 'K'DLMCGORCV-SMIRNCV CUMULATED RELAY OCCURRENCES IN:
010050 01 05 OUTPT-FIVE FREQUENCY TABLE FÜR INDEPENDENT LETTER OCCURRENCES IN:
010060 01 05 OUTPT-FOUR-KS-LANG PIC X(30).
010070 01 05 OUTPT-FOUR-G PIC X(9) VALUE SPACES.
010080 01 05 OUTPT-FILLER-PIC X(37) VALUE SPACES *LEITER*.
010090 01 05 OUTPT-FILLER-PIC X(25) VALUE SPACES *LEITER*.
010100 01 05 OUTPT-FILLER-PIC X(71) VALUE *JUM. REL. FREQ.*.
010110 01 05 OUTPT-FILLER-PIC X(40) VALUE SPACES.
010120 01 05 OUTPT-FILLER-PIC X(25) VALUE SPACES.
010130 01 05 OUTPT-FILLER-PIC 9.99999 SPACES.
010140 01 05 OUTPT-FILLER-PIC X(60) VALUE SPACES.
010150 01 05 OUTPT-FILLER-PIC X(28) VALUE *OVERFLOW ERROR AT STATEMENT*.
010160 01 05 OUTPT-FOUR-LANG PIC X(7).
010170 01 05 OUTPT-FIVE-A PIC X(49) VALUE SPACES *JOINT PROBABILITY TABLE FOR X(1)-ANAL-ONE
010180 01 05 OUTPT-FILLER-PIC X(37) VALUE SPACES.

```

C11850-
 011900 01 'X(2)' FILLER PIC X(47) VALUE SPACES.
 00120000 01 OUTPT-FIVE-B PIC X(11) VALUE SPACES.
 00120500 05 FILLER PIC X(25) VALUE SPACES.
 00121000 05 YULES-K-2 PIC Z(12)9.999 S K FOR X(1)-X(2):'.
 00121500 05 OUTPT-SIX-A PIC X(80) VALUE SPACES.
 00122000 01 OUTPT-FILLER PIC X(49) VALUE SPACES.
 00122500- 05 FILLER PIC X(37) VALUE SPACES.
 00123000- 0-X(3)- FILLER PIC X(47) VALUE SPACES.
 00124000 01 OUTPT-SIX-B PIC X(11) VALUE SPACES.
 00124500 05 FILLER PIC X(25) VALUE SPACES.
 00125000 05 YULES-K-3 PIC Z(12)9.999 S K FOR X(1)-X(3):'.
 00125500 05 FILLER PIC X(80) VALUE SPACES.
 00126000 01 CARD-FILL TEST PIC X.
 00126500 05 SEQ-NR PIC X(10).
 00127000 05 LANG-ID PIC X(10).
 00127500 05 CRF-CUT PIC 9V9(5), OCCURS 11 TIMES.
 00012850 05 PROCEDURE DIVISION.
 00012950 00130000 START-UP.
 00013000 BEGIN READ POPULATION AT END GO TO EOJ.
 00013100 IF NEW-LANG-PERFORM CLOSE-OUT-RT THRU INITIALIZE-RT.
 00013150 PERFORM DAT-SHIFT VARYING X FROM 1 BY 1 UNTIL X > 71.
 00013200 ADD 1 TO CARD-NR ON SIZE ERROR GO TO E-1.
 00013250 MOVE 3 TO N.
 00013300 DATA-CCLLCC (N) = SPACE GO TO BLANK-CCLL-STEP.
 00013350 MOVE 1 TO A.
 00013400 CK-CNE.
 00013450 IF CC (N) = LTR (A) GO TO PROC-TWO.
 00013500 ADD 1 TO CK-CNE.
 00013550 ADD 1 TO CK-CNE.
 00013600 CK-TWO: ADD 1 TO CK-CNE.
 00013650 COMPUTE M = N + 1.
 00013700 IF CC (M) = SPACE GO TO COL-STEP.
 00013750 IF CC (M) = '-' AND M = 73 PERFORM CARD-STEP-ONE.
 00013800 MOVE 1 TO B.
 00013850 CK-TWO: IF CC (M) = LTR (B) GO TO PROC-THREE.
 00013900 IF 3 = 33 GO TO COL-STEP.
 00013950
 00014000
 00014050
 00014100
 00014150
 00014200

```

014250 ADD 1 TC B.
014300 GO TO CK-TWO.
014350 PROC-THREE.
014400 ADD 1 TO FCL-FREQ (A, B) ON SIZE ERROR GO TO E-3.
014450 IF A < 11 AND B < 11 ADD 1 TO F-VOW-SIZE ERRCR GO TO E-4.
014500 IF A < 11 AND B NOT < 11 ADD 1 TO F-CONS-VOW ON SIZE
014550 IF ERROR GO TO E-5.
014600 IF A NOT < 11 AND B < 11 ADD 1 TO F-VOW-CONS ON SIZE
014650 IF A NOT < 11 AND B NOT < 11 ADD 1 TO F-CONS-CONS ON SIZE
014700 IF ERROR GO TO E-7.
014750 COMPUTE L = N + 2.
014800 IF CC (L) = SPACE GO TO COL-STEP.
014850 IF CC (L) = '-' AND L = 73 PERFORM CARD-STEP-TWO.
014900 MOVE E 1 TO B.

CK-THREE C (L) = LTR (B) PERFORM ADER.
014950 IF CC (L) = LTR (B) GO TO COL-STEP.
015000 IF B = 23 GO TO COL-STEP.
015050 ADD 1 TO B.
015100 GO TO CK-THREE.

COL-STEP. NOT < 72 GO TO BEGIN.
015150 ADD 1 TO N.
015200 GO TO DATA-COLL.
015250 EGJ.

PERFORM CLOSE-CUT-RT TIMES.
015300 MOVE SPACES TO PRINT-LINE.
015350 WRITE POPULATION AFTER TOP-CF-PAGE.
015400 CLOSE POPULATION, LANG-ANAL, CARD-FILE.
015450 STOP RUN.

LUCP-RT MOVE SPACES TO PRINT-LINE.
015500 WRITE PRINT-LINE AFTER 1.
015550 BLANK-COLL-STEP.
015600 ADD 1 TO N.
015650 IF N = 73 GO TO BEGIN.
015700 GO TO DATA-COLL.
015750 CARD-STEP-DONE.

MOVE 2 TO N.
015800 MOVE CC (72) TO CC (2).
015850 READ POPULATION AT END GO TO EOJ.
015900 PERFORM DAT-SHIFT VARYING X FROM 1 BY 1 UNTIL X > 71.
015950 ADD 1 TO CARD-NR ON SIZE ERRCR GO TC E-9.
016000 COMPUTE N = N + 1.
016050 CARD-STEP-N.
016100 MOVE 1 TO N.

```

```

016650 MOVE CC (71) TO CC (2).
016700 ADDN.PUTER L = N + 2.
016750 PERFORAT ION AT END GO TC EOJ.
016800 PERFORAT SHIFT VARYING X FROM 1 BY 1 UNTIL X > 71.
016850 CARD-NR ON SIZE ERROR GO TO E-10.
016900 CLOSE-CUTTRT. CARD-NR = 0 GO TO INITIALIZE-RT.
016950 IF C-ZER OS-TC TOTAL-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017000 PERFORM IND-PREP-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017050 PERFORM OUT-PT-GNE-PRIN-T VARYING A FROM 1 BY 1 UNTIL A > 33.
017100 PERFORM OUT-TOTAL-O ONE-F-ILF-TLT* FROM OUTPT-CNE-F BEFORE 1.
017150 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017200 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017250 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017300 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017350 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017400 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017450 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017500 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017550 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017600 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017650 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017700 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017750 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017800 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017850 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017900 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
017950 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018000 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018050 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018100 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018150 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018200 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018250 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018300 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018350 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018400 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018450 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018500 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018550 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018600 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018650 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018700 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018750 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018800 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018850 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018900 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
018950 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
019000 PERFORM OUT-PT-TW-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.

```



```

021450 PRINT-LINE FROM OUTPT-CNE-A AFTER TOP-OFF-PAGE.
021550 PRINT-LINE FROM LANG-LINE AFTER 1.
021600 PRINT-LINE FROM CUTPT-CNE-C BEFORE 2.
021650 PRINT-LINE FROM CUTPT-CNE-D BEFORE 2.
021700 PRINT-LINE PRINT(A) TO LTR-CNE.
021750 MOVE LTR(A) TO FREQ-ONE.
021850 MOVE TLP(A) TO PROB-ONE.
021950 WRITE PRINT-LINE FROM OUTPT-CNE-E BEFORE 2.

022050 DEPN-TWO-CALC.
022050 MOVE ZEROS TO FOL-FREQ-TOTAL.
022050 MOVE ZERO-SUM DEPN-TWO-SUM VARYING B FROM 1 BY 1 UNTIL B > 33.
022050 MOVE ZERO-SUM DEPN-TWO-PROB VARYING B FROM 1 BY 1 UNTIL B > 33.

022150 DEPN-TWO-CALC.
022150 MOVE ZERO-SUM FOL-FREQ-TOTAL = FOL-FREQ-TOTAL + FOL-FREQ (A, B).
022150 COMPUTE PROB(A, B) > 0 COMPUTE FOL-PROB (A, B) ROUNDED =
022200 IF FOL-FREQ(A, B) / FOL-FREQ-TOTAL, ELSE MOVE ZEROS
022200 TO FOL-FREQB(A, B).

02222550 OUTPT-TWO-SPACES TO PRINT-LINE FROM OUTPT-TWO-A AFTER TOP-OFF-PAGE.
02222550 WRITE PRINT-LINE FROM OUTPT-TWO-B BEFORE 1.
02222550 WRITE PRINT-LINE FROM LANG-LINE BEFORE 0.
02222550 WRITE PRINT-LINE FROM LANG-UNDERLINE BEFORE 2.
02222550 WRITE PRINT-LINE FROM FIRST-LTR-CUTPT BEFORE 2.

02222750 CCL-TWELVE LTR(A) TO FIRST-LTR(A).
02222750 MOVE LTR(A) TO DUMP-A.
02222750 CUTPT-TWO-OR DUMP-A VARYING A FROM 1 UNTIL A > 13.
02222750 WRITE PRINT-LINE FROM DEPN-FREQ-LINE BEFORE 1.
02222750 MOVE LTR(A) TO DUMP-B VARYING A FROM 1 UNTIL A > 13.
02222750 PERFORM PRINT-LINE FROM DEPN-FRCB-LINE BEFORE 2.

02222950 CUTPT-TWO-OR DUMP-B VARYING A FROM 14 BY 1 UNTIL A > 26.
02222950 WRITE PRINT-LINE FROM DEPN-FREQ-LINE BEFORE 1.
02222950 MOVE LTR(A) TO DUMP-C VARYING A FROM 14 BY 1 UNTIL A > 26.
02222950 CUTPT-TWO-OR DUMP-C VARYING A FROM 15 BY 1 UNTIL A > 26.
02222950 WRITE PRINT-LINE FROM DEPN-PROB-LINE BEFORE 2.

02233250 CUTPT-TWO-OR DUMP-B VARYING A FROM 27 BY 1 UNTIL A > 33.
02233250 WRITE PRINT-LINE FROM DEPN-FREQ-LINE BEFORE 1.
02233250 MOVE LTR(A) TO DUMP-B VARYING A FROM 27 BY 1 UNTIL A > 33.
02233250 PERFORM PRINT-LINE FROM DEPN-PROB-LINE BEFORE 2.

02233450 CCL-TWELVE LTR(A) TO FIRST-LTR(A).
02233450 MOVE LTR(A) TO 13.

```

023850 CCL-TITLE-C.
 MOVE LTR (A) TO FIRST-LTR (I).
 MOVE ZEROES TO FOL-FREQ-TOTAL.
 PERFORM DEPN-TWO-SUM VARYING B FROM 34 BY 1 UNTIL B > 66.
 CUTPT-TIE SPACES IN PRINT-LINE FROM CUTPT-THREE-A AFTER TOP-OF-PAGE.
 WRITE PRINT-LINE FROM CUTPT-TWO-B BEFORE 1.
 WRITE PRINT-LINE FROM LANG-LINE BEFORE 0.
 WRITE PRINT-LINE FROM LANG-UNDERLINE BEFORE 2.
 WRITE PRINT-LINE FROM FIPST-LTR-CUTPT BEFORE 2.
 VC-DUMP
 COMPUTE VC-TOTAL = F-CGNS-CONS + F-VOW-CONS + F-CONS-VOW
 + F-VOW-VOW.
 IF F-VOW-VOW > 0 COMPUTE P-VOW-VOW ROUNDED = F-VOW-VOW /
 VC-TOTAL ELSE MOVE ZEROES P-VOW-ROUNDED = F-CONS-VOW /
 F-CGNS-CONS P-CONS-VCNWD = F-CONS-CONS /
 VC-TOTAL ELSE MOVE ZEROES P-CONS-VCNWD = F-VOW-CONS /
 F-VOW-CONS
 IF F-VCTOTAL ELSE MOVE ZEROES P-CONS-VCNWD = F-VOW-CONS /
 F-CGNS-CONS P-CONS-VCNWD = F-VOW-CONS /
 VC-TOTAL ELSE MOVE ZEROES P-CONS-VCNWD = F-VOW-CONS /
 F-CGNS-CONS P-CONS-VCNWD = F-VOW-CONS /
 WRITE PRINT-LINE FROM LANG-LINE AFTER TOP-OF-PAGE.
 WRITE PRINT-LINE FROM LANG-UNDERLINE BEFORE 3.
 WRITE PRINT-LINE FROM OUTPT-FOUR-A BEFORE 2.
 WRITE PRINT-LINE FROM OUTPT-FOUR-B BEFORE 2.
 WRITE PRINT-LINE FROM OUTPT-FOUR-C BEFORE 2.
 WRITE PRINT-LINE FROM OUTPT-FOUR-D BEFORE 3.
 YULE-PREP
 COMPUTE S = S + (!LF (A) ** 2).
 KS-CALC
 COMPUTE CRF-TOTAL = CRF-TOTAL + ILP (A).
 MOVE LTR (A) TO KS-LTR.
 MOVE CRF-TOTAL TO KS-CRF.
 WRITE PRINT-LINE FROM KS-LINE BEFORE 1.
 E-1. MOVE 014300 TO ER-STEP.
 GO TO DUMP.
 E-2. MOVE 014901 TO ER-STEP.
 GO TO DUMP.
 E-3. MOVE 015400 TO ER-STEP.
 GO TO DUMP.
 E-4. MOVE 026200

026250 MOVE 015450 TO ER-STEP.
 GO TO DUMP.
 E-5. MOVE 015550 TO ER-STEP.
 GO TO DUMP.
 E-6. MOVE 015650 TO ER-STEP.
 GO TO DUMP.
 E-7. MOVE 015750 TO ER-STEP.
 GO TO DUMP.
 E-8. MOVE 016000 TO ER-STEP.
 GO TO DUMP.
 E-9. MOVE 017300 TO ER-STEP.
 GO TO DUMP.
 E-10. MOVE 017500 TO ER-STEP.
 DUMP. MOVE CARD-NR TIC ER-CARD FROM ERROR-LINE AFTER 2.
 WRITE PRINT-LINE FROM ERROR-LINE.
 GO TO EOJ.
 DUMP-A. IF B > 33 COMPUTE L = B - 33, MOVE LTR (L) TO FREQ-ROW-LTR,
 ELSE MOVE LTR (B) TO FREQ-ROW-LTR,
 IF A < 14 MOVE FCL-FREQ (A) TO DEPN-FREQ (A),
 IF A > 13 AND A < 27 COMPUTE M = A - 13, MOVE FCL-FREQ (A, B),
 IF A > 26 COMPUTE M = A - 26, MOVE FCL-FREQ (A, B),
 TO DEPN-FREQ (M).
 DUMP-B. IF B > 33 COMPUTE L = B - 33, MOVE LTR (L) TO PROB-ROW-LTR,
 ELSE MOVE LTR (B) TO PROB-ROW-LTR,
 IF A < 14 MOVE FCL-PROB (A) TO DEPN-PROB (A),
 IF A > 13 AND A < 27 COMPUTE M = A - 13, MOVE FCL-PROB (A, B),
 IF A > 26 COMPUTE M = A - 26, MOVE FOL-PROB (A, B),
 TO DEPN-PROB (M).
 02627450
 02627500
 02627550
 02627600
 02627750
 026277650
 026277750
 026277800
 02627900
 02627950
 02628000
 02628050
 02628100
 02628150
 02628200
 02628250
 02628300
 02628350
 02628400
 02628450
 02628500
 02628550
 02628600

MOVE ZEROES TO ILF (A).
 FLUSHING.
 MOVE SPACES TO FIRST-LTR (A).
 MOVE ZEROS TO DEPN-FREQ (A).
 MOVE ZEROS TO DEPN-PROB (A).
 MOVE ZEROS TO FOL-FREQ (A, B).

```

028650 MOVE ZEROS TO FOL-PROB (A, B).
028750 ADER COMPUTE B = B + 33
028800 ADD 1 TO FOL-FREQ (A, B) ON SIZE ERROR GO TO E-8.
028850 COMPUTE B = B - 33.
02890000 DAT-
02895000 SHIFT-PUTE Y = X + 2 * (Y).
02910000 JOIN MOVE PROB-ZEROS TO FOL-FREQ-TOTAL.
02915000 PERFORM B-EROM 1 BY 1 UNTIL B > 33.
02920000 PERFORM DEPN-TWO-PKCS VARYING A FROM 1 BY 1 UNTIL A > 33.
02925000 AFTER B-FROM 1 BY 1 UNTIL B > 33.
02930000 PERFORM OUT-PT-FIVE-PREP.
02935000 PERFORM OUTL-TITLE-A VARYING A FROM 1 BY 1 UNTIL A > 13.
02940000 WRITE PRINT-LINE FROM FIRING-LTR-TITLE BEFORE 2.
02945000 PERFORM OUT-PT-DUMP-P-U VARYING B FROM 1 BY 1 UNTIL B > 33.
02950000 PERFORM OUT-PT-FIVE-PREP.
02955000 PERFORM CCL-TITLE-B VARYING A FROM 1 BY 1 UNTIL A > 26.
02960000 WRITE PRINT-LINE FROM FIRING-B BEFORE 2.
02965000 PERFORM CUT-PT-DUMP-F-E VARYING B FROM 1 BY 1 UNTIL B > 33.
02970000 PERFORM FLUSHING-VARYING A FROM 1 BY 1 UNTIL A > 13.
02975000 PERFORM CNT-LINE FROM FIRING-LTR-TITLE BEFORE 2.
02980000 PERFORM OUT-PT-FIVE-PREP.
02985000 PERFORM CNT-LINE FROM FIRING-B BEFORE 2.
02990000 MOVE ZEROS TO S-YULES-K-PREP.
02995000 MOVE ZEROS TO S-YULES-K-2 ROUNDED = (10000 * (S - FOL-FREQ-TOTAL)) /
03000000 PERFORMANCE YULES-K-2 TEST.
03005000 MOVE ZEROS TO FOL-FREQ-TOTAL #2.
03010000 WRITE PRINT-LINE FROM OUT-PT-FIVE-B BEFORE 1.
03015000 MOVE 0-B.
03020000 MOVE TO TEST.
03025000 PERFORM CARD-DUMP VARYING A FROM 1 BY 1 UNTIL A > 33.
03030000 MOVE ZEROS TO FOL-FREQ-TOTAL.
03035000 PERFORM DEPN-TNO-SUN VARYING A FROM 1 BY 1 UNTIL A > 33.
03040000 PERFORMANCE B-FROM 34 BY 1 UNTIL B > 66.
03045000 PERFORMANCE DEPN-TWO-PROB VARYING A FROM 34 BY 1 UNTIL A > 33.
03050000 AFTER B-FROM 34 BY 1 UNTIL B > 66.
03055000 PERFORM OUT-PT-SIX-PREP.
03060000 PERFORM CCL-TITLE-A VARYING A FROM 1 BY 1 UNTIL A > 13.
03065000 WRITE PRINT-LINE FROM FIRING-LTR-TITLE BEFORE 2.
03070000 PERFORM OUT-PT-DUMP-P-U VARYING B FROM 34 BY 1 UNTIL B > 66.
03075000 PERFORM OUT-PT-SIX-PREP.
03080000 PERFORM CCL-TITLE-B VARYING A FROM 14 BY 1 UNTIL A > 26.
03085000
03090000
03095000
03100000

```

```

031050 WRITE PRINT-LINE FROM FIRST-LTR-TITLES BEFORE 2*
031100 OUTPT-DUMP-E VARYING B FROM 34 BY 1 UNTIL B > 66.
031150
031200
031250
031300
031350
031400
031450
031500
031550
031600
031650
031700
031750 MOVE ZEROES TO S-K-PREP VARYING A FROM 1 BY 1 UNTIL A > 33
AFTER B FROM 34 BY 1 UNTIL B > 66 # (S - FCL-FREQ-TOTAL) /
COMPUTE YULES-K-3 ROUNDED = (10000**2)**2
{FOL-FREQ-TOTAL-FOL-LINE-FROM OUTPT-SIX-B BEFORE 1.
MOVE ZEROS TO CTEST.
MOVE CARD-DUMP VARYING A FROM 1 BY 1 UNTIL A > 33
AFTER B FROM 34 BY 1 UNTIL B > 66.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM OUTPT-FIVE-A AFTER TOP-OFF-PAGE.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM OUTPT-TWO-THREE-B BEFORE 1.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM LANG-LINE BEFORE 0.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM LANG-UNDERLINE BEFORE 2.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM FIRST-LTR-OUTPT BEFORE 2.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM OUTPT-FIVE-A AFTER TOP-OFF-PAGE.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM OUTPT-TWO-THREE-B BEFORE 1.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM LANG-LINE BEFORE 0.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM LANG-UNDERLINE BEFORE 2.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM OUTPT-SIX-A AFTER TOP-OFF-PAGE.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM OUTPT-TWO-THREE-B BEFORE 1.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM LANG-LINE BEFORE 0.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM LANG-UNDERLINE BEFORE 2.
CUTPT-FIVE-PREP* WRITE PRINT-LINE FROM FIRST-LTR-OUTPT BEFORE 2.
MOVE YULES-K-PREP COMPUTE S = S + (FOL-FREQ (A, B) ** 2).
CARD-DUMP COMPUTE CRF-TOTAL = CRF-OUT (L).
ADD 1 TO CRF-TOTAL TO CRF-OUT (L).
MOVE WS-SEQ-NR TO WS-SEQ-NR IF L = 11 MOVE WS-SEQ-NR TO WS-SEQ-NR MOVE ZEROS TO L WRITE CARD-CUT FROM CARD-FILL.

```

APPENDIX H. POPULATION ANALYSIS PROGRAM (V-2)

```

000050 IDENTIFICATION DIVISION. V2-DIAGNOSTIC-ANALYSIS.
000100 PROGRAM-ID. PAPRUS.
000150 AUTHOR. R. A. RAU, USE.
000200 REMARKS. THIS PROGRAM IS PART OF MY MASTER OF SCIENCE IN
000250 MATHEMATICAL COMPUTERS AND IS DESIGNED TO CALCULATE VARIOUS
000300 STATISTICAL CHARACTERISTICS FOR LETTERS APPEARING IN
000350 LANGUAGES WRITTEN IN THE MODERN ALPHABET. THE SPACE
000400 EXPANSION OF THESE LANGUAGES AND MODELS SIGNIFICANTLY INCLUDE THE INTERDEPENDENCY
000450 OF LETTERS IN THESE LANGUAGES AND WORDS. INCLUDE THE INTERDEPENDENCY
000500 OF LETTERS IN THESE LANGUAGES AND WORDS.
000550 ENVIRONMENT DIVISION.
000600 CONFIGURATION-COMPUTER. IBM-360-67.
000650 SOURCE-COMPUTER. IBM-360-67.
000700 SPECIAL-NAMES-OP-CF-PAGE.
000750 FILE-CONTROL-INPUT-SECTION.
000800 INPUT-CONTROL-POPULATION ASSIGN TO UR-S-IN1.
000850 FILE-SELECT-POPULATION ASSIGN TO UR-S-OUT1.
000900 SELECT-LANG-FILE ASSIGN TO UR-S-OUT2.
000950 DATA DIVISION.
001000 FD POPULATION
001050 LABEL RECORDS ARE OMITTED
001100 BLOCK RECORDS CONTAINS 5 RECORDS
001150 DATA RECORD IS POP-CARD, LANG-CARD.
001200 01 POP-CARD
001250 05 SAMPLE-FLD.
001300 10 FILLER
001350 05 FILLER
001400 05 FILLER
001450 05 FILLER-FLD.
001500 01 LANG-CARD-TYPE 'L'.
001550 05 FILLER-NEW-LANG VALUE 'L'.
001600 05 FILLER
001650 05 LANGUAGE
001700 05 LANGUAGE
001750 05 FILLER
001800 05 FILLER-KEY
001850 05 FILLER
001900 05 LANG-KEY
001950 FD LANG-CARD
002000 LABEL RECORDS ARE OMITTED
002050 BLOCK RECORDS CONTAINS 5 RECORDS
002100 DATA RECORD IS PRINT-LINE.
002150 PRINT-LINE.
002200 01 PIC X(133).

```


014250 BRAVO. MOVE 3 TO N.
 014350 IF CK-(N) = SPACE GO TO CK-CNE.
 014400 PERFORM FIRST-LETTER-CHECK THRU FIRST-EXIT.
 014450 IF MATCH = ZERO GO TO CK-ONE-BRAVO.
 014500 ADD 1 TO ILF (A) ON SIZE ERROR GO TO E-1.
 014550 COMPUTE M = N + 1. SHIFT-TWO.
 014600 IF CL-(M) = 73 PERFORM GOTO CK-TWO.
 014650 PERFORM SECOND-LETTER-CHECK THRU SECOND-EXIT.
 014700 IF MATCH = ZERO GO TO CK-TWO.
 014750 CHARLIE. ADD 1 TO FOL-FREQ (A, B) ON SIZE ERROR GO TO E-2.
 014800 PERFORM VC-TALLEY THRU VC-EXIT.
 014850 COMPUTE L = N + 2. IF L = 73 AND CL-(73) = 'I' PERFORM SHIFT-THREE.
 014900 IF L = 73 AND CL-(73) NOT = 'I' PERFORM ALT-THREE.
 014950 IF CL-(L) = SPACE GO TO CK-THREE.
 015000 PERFORM THIRD-LETTER-CHECK THRU THIRD-EXIT.
 015050 IF MATCH = ZERO GO TO CK-THREE.
 015100 DELTA. COMPUTE B = B + 34.
 015150 ADD 1 TO FUL-FREQ (A, B) ON SIZE ERROR GO TO E-3.
 015200 ECHO. ADD 1 TO N.
 015250 GO TO BRAVO.
 015300 CK-ONE. ADD 1 TO ILF (34) ON SIZE ERROR GO TO E-4.
 015350 ADD 1 TO N.
 015400 MOVE 34 TO A.
 015450 CK-ONE-ALPHA. IF N = 73 PERFORM SHIFT-ONE.
 015500 IF CL-(N) = SPACE GO TO CK-CNE-BRAVO.
 015550 COMPUTE N = N - 1.
 015600 PERFORM SECOND-LETTER-CHECK THRU SECOND-EXIT.
 015650 IF MATCH = 1 GO TO CHARLIE, ELSE ADD 1 TO N.
 015700 CK-CNE-BRAVO. ADD 1 TO N.
 015750 GO TO CK-CNE-ALPHA.
 015800 CK-TWO. ADD 1 TO FOL-FREQ (A, 34) ON SIZE ERROR GO TO E-5.
 015850 ADD 1 TO N.
 015900 CK-TWO-ALPHA. COMPUTE M = N + 1. SHIFT-TWO.
 015950 IF CL-(M) = 73 PERFORM SPACE GO TO CK-TWO-BRAVO.
 016000 COMPUTE N = N - 1.
 016050 PERFORM THIRD-LETTER-CHECK THRU THIRD-EXIT.
 016100
 016150
 016200
 016250
 016300
 016350
 016400
 016450
 016500
 016550
 016600

016650 CK-TWO-BRAVO. N.
 0016750 ADD 1 TO CK-TWO-ALPHA.
 0016800 CK-THREE * ADD 1 TO FOL-FREQ (A, 68) ON SIZE ERROR GO TO E-6.
 0016900 GO TO ECHO.
 0017000 SHIFT ONE CL (72) TO CL (2).
 0017050 MOVE SPACES TO CL (1), CL (2).
 0017100 MOVE 3 TC N.
 0017150 SHIFT PUFFEM READ-A-CARD.
 0017200 MOVE 3 TC N.
 0017250 MOVE CL (72) TO CL (2).
 0017300 MOVE FORM READ-A-CARD.
 0017350 MOVE 2 TC N.
 0017400 MOVE 3 TO M.
 0017450 SHIFT MOVE CL (72) TO CL (2).
 0017500 MOVE FORM READ-A-CARD.
 0017550 MOVE 1 TO N.
 0017600 MOVE 3 TO L.
 0017650 ALT-MOVE CL (72) TO CL (1).
 0017700 MOVE SPACES TC CL (2).
 0017750 PERFORM READ-A-CARD.
 0017800 MOVE 0 TC N.
 0017850 MOVE 2 TC L.
 0017900 READ-A-CARD READ POPULATION AT END GO TO EOJ.
 0017950 IF NEW-LANG GO TO FLUSH.
 0018000 ADD 1 TO CARD-NR ON SIZE ERROR GO TO E-7 UNTIL X > 71.
 0018050 PERFORM DAT-SHIFT VARYING X FROM 1 BY 1 UNTIL X > 71.
 0018100 COMPUTE Y = X + 2.
 0018150 MOVE CC (X) TO CL (Y).
 0018200 FLUSH.
 0018250 IF CARD-NR NOT = ZERO PERFORM CLOSE-OUT-RT.
 0018300 GO TO ALPHA.
 0018350 EOJ.
 0018400 PERFORM CLOSE-CUT-RT TIMES.
 0018450 MOVE SPACES TC PRINT-LINE.
 0018500 WRITE PRINT-LINE AFTER TOP-OFF-PAGE.
 0018550 CLOSE POPULATION, LANG-ANAL, CARU-FILE.
 0018600 CUTUP RUN.
 0018650 LCOP-RT.
 0018700 MOVE SPACES TC PRINT-LINE.
 0018750
 0018800
 0018850
 0018900
 0018950
 0019000

```

019050 WRITE PRINT-LINE AFTER 1.
019100 FIRST-LETTER-CHECK.
019150 MOVE 1 TO A.
019200 MOVE 2 ZERO TO B. MATCH.
019250 FLC-ALPHA. IF CL (N) = LTR (A) GO TO FLC-BRAVO.
019300 IF A = 33 GO TO FIRST-EXIT.
019350 ADD 1 TO A.
019400 GO TO FLC-ALPHA.
019450 FLC-BRAVO. MOVE 1 TO MATCH.
019500 IF CL (N) = LTR (B) GO TO SLC-BRAVO.
019550 FIRST-EXIT. EXIT.
019600 SECOND-LETTER-CHECK.
019650 MOVE 1 TO B. MATCH.
019700 MOVE 2 ZERO TO N + 1.
019750 COMPUTE N = N + 1.
019800 SLC-ALPHA. IF CL (M) = LTR (B) GO TO SECOND-EXIT.
019850 MOVE 1 TO B.
019900 ADD 1 TO B.
019950 GO TO SLC-ALPHA.
020000 SLC-BRAVO. MOVE 1 TO MATCH.
020050 SECOND-EXIT. EXIT.
020100 THIRD-LETTER-CHECK.
020150 MOVE 1 TO B.
020200 MOVE 2 ZERO TO N + 2.
020250 MOVE 1 TO B. MATCH.
020300 MOVE 1 TO B.
020350 MOVE 2 ZERO TO N + 2.
020400 COMPUTE L = N + 2.
020450 TLC-ALPHA. IF CL (L) = LTR (B) GC TO TLC-BRAVO.
020500 IF B = 33 GO TO THIRD-EXIT.
020550 ADD 1 TO B.
020600 GO TO TLC-ALPHA.
020650 SLC-BRAVO. MOVE 1 TO MATCH.
020700 MOVE 1 TO B.
020750 MOVE 1 TO B. EXIT.
020800 THIRD-EXIT. EXIT.
020850 VC-TALLEY. IF L = 34 GO TO VC-EXIT.
020900 IF A < 11 AND B < 11 ADD 1 TO F-
020950 IF A < 11 AND B NOT < 11 ADD 1 TO F-
021000 IF A ERROR GO TO E-9.
021050 IF A NOT < 11 AND B < 11 ADD 1 TO F-
021100 IF A ERROR GO TO E-10.
021150 IF A NOT < 11 AND B < 11 ADD 1 TO F-
021200 IF A ERROR GO TO E-11.
021250 IF A NOT < 11 AND B < 11 ADD 1 TO F-
021300 VC-EXIT. EXIT.
021350 INITIALIZE-FERDS TO VC-TOTAL, A, B, L, M,

```

021450
0021550
0021600
0021650
0021700
0021750
0021800
0021850
0021900
0021950
0022000
0022050
0022100
0022150
0022200
0022250
0022300
0022350
0022400
0022450
0022500
0022550
0022600
0022650
0022700
0022750
0022800
0022850
0022900
0022950
0023000
0023050
0023100
0023150
0023200
0023250
0023300
0023350
0023400
0023450
0023500
0023550
0023600
0023650
0023700
0023750
0023800

F-YOW-CONS, F-CONST-VOW, F-VOW-VOW, FOL-FREQ-TOTAL, S,
FLF-CTOTAL, CRF-TOTAL, CARD-NK, WS-SEQ-NR.
PERFORM ZEROING-ONE VARYING A FROM 1 BY 1 UNTIL A > 34.
PERFORM ONE VARYING A FROM 1 BY 1 UNTIL A > 34.
MOVE LANG-GER TO LANG-DESIG, OUTPT-FCUR-LANG,
MOVE LANG-KEY TO LANG-ID, L-LANG-ID.
CLOSE SUBTRACT 1 FROM TLF (34).
SUBTRACT IND-TCALC VARYING A FROM 1 BY 1 UNTIL A > 34.
SUBTRACT IND-PRCNE-CALC VARYING A FROM 1 BY 1 UNTIL A > 34.
SUBTRACT OUTPT-PRNT VARYING A FROM 1 BY 1 UNTIL A > 34.
SUBTRACT OUTPT-ONE-F-ILF-TLT ONE-F-OUTPT-CNE-F BEFORE 1
SUBTRACT OUTPT-DEPN-TWOPREP. VARYING A FROM 1 BY 1 UNTIL A > 34.
SUBTRACT OUTPT-TITLE-S-A VARYING A FROM 1 BY 1 UNTIL A > 13.
SUBTRACT OUTPT-FIRST-LTR-TLES BEFORE 2 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-A VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-B VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-S-B VARYING A FROM 14 BY 1 UNTIL A > 26.
SUBTRACT OUTPT-FIRST-LTR-TLES BEFORE 2 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-B VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-C VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-D VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-E VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-F VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-G VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-H VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-I VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-J VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-K VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-L VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-M VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-N VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-O VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-P VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-Q VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-R VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-S VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-T VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-U VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-V VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-W VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-X VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-Y VARYING B FRGM 1 BY 1 UNTIL B > 23.
SUBTRACT OUTPT-TITLE-NO-DUMP-Z VARYING B FRGM 1 BY 1 UNTIL B > 23.


```

0262500 MOVE ZEROS TO S. PREP VARYING A FROM 1 BY 1 UNTIL A > 34
0263000 AFTER B FROM 1 BY 1 UNTIL B > 34. *(S - FCL-FREQ-TOTAL) /
0263500 COMPUTE YULES-K-2 ROUNDED = (10000 # (S - FCL-FREQ-TOTAL)) /
0264000 (FOL-FREQ-TOTAL)**2. *TEST*
0264500 WRITETO TEST, L-TEST, WS-SEQ-NR, L.
0265000 MOVE ZEROS TO TEST, L-TEST, WS-SEQ-NR, L.
0265500 PERFORM CARD-DUMP VARYING A FROM 1 BY 1 UNTIL A > 34
0266000 AFTER B FROM 1 BY 1 UNTIL B > 34.
0266500 MOVE ZEROS TO FOL-FREQ-TOTAL.
0267000 PERFORM DEPN-TWO-SUM VARYING A FROM 1 BY 1 UNTIL A > 34
0267500 AFTER B FROM 35 BY 1 UNTIL B > 68.
0268000 PERFORM DEPN-TWO-PROB VARYING A FROM 1 BY 1 UNTIL A > 34
0268500 AFTER B FROM 35 BY 1 UNTIL B > 68.
0269000 PERFORM CUT-PT-SIX-PREP VARYING A FROM 1 BY 1 UNTIL A > 34.
0269500 PERFORM PRINT-LINE FROM FIRST-LTR-TITLE BEFORE 2.
0270000 PERFORM CUT-PT-DUMP PT-SIX-PREP VARYING B FROM 35 BY 1 UNTIL B > 68.
0270500 PERFORM COL-TITLE-S-E VARYING A FROM 1 BY 1 UNTIL A > 34.
0271000 PERFORM COL-TITLE-S-E VARYING B FROM 35 BY 1 UNTIL B > 68.
0271500 PERFORM P9INUT-PT-DUMP PT-SIX-PREP VARYING B FROM 35 BY 1 UNTIL B > 68.
0272000 PERFORM OUT-PT-DUMP PT-SIX-PREP VARYING B FROM 35 BY 1 UNTIL B > 68.
0272500 PERFORM FLUSH-TITLE-S-C VARYING A FROM 1 BY 1 UNTIL A > 34.
0273000 PERFORM PRINT-LINE FROM FIRST-LTR-TITLE BEFORE 2.
0273500 PERFORM P9INUT-PT-DUMP PT-SIX-PREP VARYING B FROM 35 BY 1 UNTIL B > 68.
0274000 MOVE ZEROS TO S. PREP VARYING A FROM 1 BY 1 UNTIL A > 34
0274500 AFTER B FROM 35 BY 1 UNTIL B > 68. *(S - FCL-FREQ-TOTAL) /
0275000 COMPUTE YULES-K-3 ROUNDED = (10000 # (S - FCL-FREQ-TOTAL)) /
0275500 (FOL-FREQ-TOTAL)**2. *TEST*
0276000 WRITETO TEST, L-TEST, WS-SEQ-NR, L.
0276500 MOVE ZEROS TO TEST, L-TEST, WS-SEQ-NR, L.
0277000 PERFORM CARD-DUMP VARYING A FROM 1 BY 1 UNTIL A > 34
0277500 AFTER B FROM 35 BY 1 UNTIL B > 68.
0278000 IND-TCALC-CALC-F-TOTAL = ILF-TOTAL + ILF (A).
0278500 IND-PPCA-CALC-CALC
0279000 IF ILF (A) > 0 COMPUTE ILP (A) ROUNDED =
0279500 DEPN-TWO-ZEROS TO FOL-FREQ-TOTAL.
0280000 MOVE ZEROS TO DEPN-TWO-SUM VARYING B FROM 1 BY 1 UNTIL B > 34.
0280500 PERFORM DEPN-TWO-PROB VARYING B FROM 1 BY 1 UNTIL B > 34.
0281000 DEPN-TWO-PROB-CALC

```

```

MOVE ZEROS TO FOL-FREQ-TOTAL.
PERFORM DEPN-TWO-PROB VARYING B FROM 35 BY 1 UNTIL B > 68.
DEPN-TWO-SUM = FCL-FREQ-TOTAL + FOL-FREQ-TOTAL.
IF FOL-FREQ-(A,B) > 0 COMPUTE FOL-PROG-(A,B) ROUNDED =
    TO FOL-FREQ-(A,B).
    MOVE FOL-FOL-PROB-(A,B) TO FOL-FREQ-TOTAL, ELSE MOVE ZEROS
    TO PREP.
    MOVE ONE-SPACES TO PRINT-LINE FROM OUTPT-CNE-A AFTER TOP-OFF-PAGE.
    WRITE PRINT-LINE FROM OUTPT-CNE-B BEFORE 1.
    WRITE PRINT-LINE FROM LANG-LINE AFTER 1.
    WRITE PRINT-LINE FROM OUTPT-CNE-C AFTER 2.
    WRITE PRINT-LINE FROM OUTPT-CNE-D BEFORE 2.
    OUTPT ONE-PRINT :.
    MOVE LTR-CNE-(4) TO LTR-CNE-E.
    MOVE LTR-CNE-(4) TO FREQ-CNE-E.
    MOVE LTR-CNE-(4) TO PROB-CNE-E.
    MOVE LTR-CNE-(4) TO PRINT-LINE FROM OUTPT-CNE-E BEFORE 2.
    OUTPT TWO-SPACES TO PRINT-LINE FROM OUTPT-TWO-A AFTER TOP-OFF-PAGE.
    MOVE PRINT-LINE FROM OUTPT-TWO-B BEFORE 1.
    MOVE PRINT-LINE FROM OUTPT-TWO-C BEFORE 2.
    MOVE PRINT-LINE FROM OUTPT-TWO-D BEFORE 3.
    MOVE PRINT-LINE FROM OUTPT-TWO-E BEFORE 4.
    MOVE PRINT-LINE FROM OUTPT-TWO-F BEFORE 5.
    MOVE PRINT-LINE FROM OUTPT-TWO-G BEFORE 6.
    MOVE PRINT-LINE FROM OUTPT-TWO-H BEFORE 7.
    MOVE PRINT-LINE FROM OUTPT-TWO-I BEFORE 8.
    MOVE PRINT-LINE FROM OUTPT-TWO-J BEFORE 9.
    MOVE PRINT-LINE FROM OUTPT-TWO-K BEFORE 10.
    MOVE PRINT-LINE FROM OUTPT-TWO-L BEFORE 11.
    MOVE PRINT-LINE FROM OUTPT-TWO-M BEFORE 12.
    MOVE PRINT-LINE FROM OUTPT-TWO-N BEFORE 13.
    MOVE PRINT-LINE FROM OUTPT-TWO-O BEFORE 14.
    MOVE PRINT-LINE FROM OUTPT-TWO-P BEFORE 15.
    MOVE PRINT-LINE FROM OUTPT-TWO-Q BEFORE 16.
    MOVE PRINT-LINE FROM OUTPT-TWO-R BEFORE 17.
    MOVE PRINT-LINE FROM OUTPT-TWO-S BEFORE 18.
    MOVE PRINT-LINE FROM OUTPT-TWO-T BEFORE 19.
    MOVE PRINT-LINE FROM OUTPT-TWO-U BEFORE 20.
    MOVE PRINT-LINE FROM OUTPT-TWO-V BEFORE 21.
    MOVE PRINT-LINE FROM OUTPT-TWO-W BEFORE 22.
    MOVE PRINT-LINE FROM OUTPT-TWO-X BEFORE 23.
    MOVE PRINT-LINE FROM OUTPT-TWO-Y BEFORE 24.
    MOVE PRINT-LINE FROM OUTPT-TWO-Z BEFORE 25.
    CUL-MAIN-LTR-(A) TO FIRST-LTR-(A).
    OUTPT DUMP-A VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-B VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-C VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-D VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-E VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-F VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-G VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-H VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-I VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-J VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-K VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-L VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-M VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-N VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-O VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-P VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-Q VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-R VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-S VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-T VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-U VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-V VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-W VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-X VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-Y VARYING A FROM 1 BY 1 UNTIL A > 13.
    PERFORM DUMP-Z VARYING A FROM 1 BY 1 UNTIL A > 13.
    CCL-TITLESE-B COMPUTE L = A - 13.

```

```

031100 COL-T!LTER (A) TO FIRST-LTR (L).
031150 MOVE LTR (A) TO 26 FIRST-LTR (L).
031200 CUTPT THREE SPACES TO PRINT-LINE FROM OUTPT-THREE-A AFTER TOP-OFF-PAGE.
031250 WRITE PRINT-LINE FROM OUTPT-TWO-THREE-B BEFORE 1.
031300 WRITE PRINT-LINE FROM LANG-LINE BEFORE 0.
031350 WRITE PRINT-LINE FROM LANG-UNDERLINE BEFORE 2.
031400 WRITE PRINT-LINE FROM FIRST-LTR-DUTPT BEFORE 2.
031450 WRITE PRINT-LINE FROM VC-DUMP COMPUTE VC-TOTAL = F-CONS-CONS + F-VCW-CONS + F-VOW-VOW
031500 + F-CONS-VOW. COMPUTE P-YCN-VOW ROUNDED = F-VOW-VOW /
031550 0 IF F-VOW-VAL ELSE MOVE ZERO P-VCW-VCW-ROUNDED = F-CONS-VCW /
031600 0 IF F-CONS-TVAL ELSE MOVE ZERO P-CNS-VCW-ROUNDED = F-VOW-CCNS /
031650 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-CNS-VCW-ROUNDED = F-VOW-CCNS /
031700 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
031750 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
031800 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
031850 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
031900 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
031950 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
032000 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
032050 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
032100 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
032150 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
032200 0 IF F-VCH-CNS > 0 COMPUTE CS-TC P-VCH-CNS-VCW-ROUNDED = F-VOW-CCNS /
032250 WRITE PRINT-LINE FROM OUTPT-FOUR-1A BEFORE 2.
032300 WRITE PRINT-LINE FROM OUTPT-FOUR-1B BEFORE 2.
032350 WRITE PRINT-LINE FROM OUTPT-FOUR-C BEFORE 2.
032400 WRITE PRINT-LINE FROM OUTPT-FOUR-D BEFORE 2.
032450 YULE-PPREP COMPUTE S = S + (ILF (A) ** 2).
032500 KS-CALC COMPUTE CRF-TOTAL = CRF-TOTAL + YLP (A).
032550 MOVE LTR (A) TO KS-LTR.
032600 MOVE CRF-TOTAL TO KS-CRF.
032650 WRITE PRINT-LINE FROM KS-LINE BEFORE 1.
032700 DUMP-A.
032750 IF B > 34 COMPUTE L = B - 34, MCVE LTR (L) TO FREQ-ROW-LTR.
032800 ELSE MOVE LTR (B) TO FREQ-ROW-LTR.
032850 IF A < 14 MOVE FDL-FREQ (A,E) TO DEPN-FREQ (A,E).
032900 IF A > 13 AND A < 27 COMPUTE M = A - 13, MOVE FCL-FREQ (A,B) TO DEPN-FREQ (A,B).
032950 IF A TO DEPN-FREQ (M) = A - 26, MOVE FCL-FREQ (A,B).
033000 IF A > 26 COMPUTE M = A - 26, MOVE FCL-FREQ (A,B).
033050 DUMP-B.
033100 IF B > 34 COMPUTE L = B - 34, MCVE LTR (L) TO PROB-ROW-LTR.
033150 ELSE MOVE LTR (B) TO PROB-ROW-LTR.
033200 IF A < 14 MOVE FDL-PROB (A,E) TO DEPN-PROB (A,E).
033250 IF A > 13 AND A < 27 COMPUTE M = A - 13, MOVE FCL-PROB (A,B) TO DEPN-PROB (A,B).

```

```

033450      TO DEPN-PRCB (M) = A - 26, MOVE FCL-PRCB (A, B)
033500      IF A > 26 COMPUTE (M).
033550      TO DEPN-PROB (M).
033600      ZEROING-ONE.
033650      MOVE ZEROES TO ILF (A).
033700      MOVE ZEROES TO ILP (A).
033750      FLUSHING. SPACES TO FIRST-LTR (A), DEPN-PRCB (A).
033800      MOVE TWO-ZEROS TO DEPN-FREQ (A, B).
033850      MOVE ZEROES TO FOL-PROB (A, B).
033900      MOVE ZEROES TO FOL-FREQ (A, B).
033950      CUTPT-FIVE-PRINT-LINE FROM CUTPT-FIVE-A AFTER TOP-OFF-PAGE.
034000      WRITE PRINT-LINE FROM CUTPT-TWO-THREE-E BEFORE 1.
034050      WRITE PRINT-LINE FROM LANG-LINE BEFORE 0.
034100      WRITE PRINT-LINE FROM LANG-UNDERLINE BEFORE 2.
034150      WRITE PRINT-LINE FROM FIRST-ULTR-CUTPT BEFORE 2.
034200      CUTPT-DUMP-P-D.
034250      PERFORM DUMP-B VARYING A FROM 1 BY 1 UNTIL A > 13.
034300      WRITE PRINT-LINE FROM DEPN-PROB-LINE BEFORE 2.
034350      CUTPT-DUMP-P-E.
034400      PERFORM DUMP-B VARYING A FROM 14 BY 1 UNTIL A > 26.
034450      WRITE PRINT-LINE FROM DEPN-PROB-LINE BEFORE 2.
034500      CUTPT-DUMP-P-F.
034550      PERFORM DUMP-B VARYING A FROM 27 BY 1 UNTIL A > 34.
034600      WRITE PRINT-LINE FROM DEPN-PROB-LINE BEFORE 2.
034650      CUTPT-SIX-PRINT-LINE FROM CUTPT-SIX-A AFTER TOP-OFF-PAGE.
034700      WRITE PRINT-LINE FROM CUTPT-TWO-THREE-B BEFORE 1.
034750      WRITE PRINT-LINE FROM LANG-LINE BEFORE 0.
034800      WRITE PRINT-LINE FROM LANG-UNDERLINE BEFORE 2.
034850      WRITE PRINT-LINE FROM FIRST-ULTR-CUTPT BEFORE 2.
034900      WRITE PRINT-LINE FROM YULES-K-REP.
034950      COMPUTE S = S + (FCL-FREQ (A, B) * # 2).
035000      CARD-DUMP COMPUTE CRF-TOTAL = CRF-TOTAL + FCL-PRCB (A, B).
035050      ADD 1 TO L.
035100      MOVE CRF-TOTAL TO CRF-OUT (L).
035150      IF WS-SEQ-NR = 105, ADD 1 TO WS-SEQ-NR.
035200      MOVE WS-SEQ-NR TO L-SEQ-NR.
035250      MOVE CRF-TOTAL TO L-CRF-OUT.
035300      MOVE CRF-TOTAL FROM LAST-CARD-FILL.
035350      IF L = 11 ADD 1 TO WS-SEQ-NR.
035400      MOVE WS-SEQ-NR TO SEQ-NR.
035450      MOVE ZEROS TO L.
035500      WRITE CARD-OUT FROM CARD-FILL.
035550
035600
035650
035700
035750
035800

```

035850 MOVE 016450 TO ER-STEP.
 035900 GO TO DUMP.
 E-2. MOVE 016850 TO ER-STEP.
 GO TO DUMP.
 E-3. MOVE 017300 TO ER-STEP.
 GO TO DUMP.
 E-4. MOVE 017550 TO ER-STEP.
 GO TO DUMP.
 E-5. MOVE 018200 TO ER-STEP.
 GO TO DUMP.
 E-6. MOVE 018850 TO ER-STEP.
 GO TO DUMP.
 E-7. MOVE 019800 TO ER-STEP.
 GO TO DUMP.
 E-8. MOVE 022450 TO ER-STEP.
 GO TO DUMP.
 E-9. MOVE 022500 TO ER-STEP.
 GO TO DUMP.
 E-10. MOVE 022600 TO ER-STEP.
 GO TO DUMP.
 E-11. MOVE 022700 TO ER-STEP.
 GO TO DUMP.
 DUMP MOVE CARD-NR TO ER-CARD
 WRITE PRINT-LINE FROM ERROR-LINE AFTER 2.
 GO TO EOJ.
 037500
 037550
 037600

APPENDIX I. LANGUAGE IDENTIFICATION PROGRAM (V-1)

IDENTIFICATION DIVISION-ID-VERSION-ONE.
 PROGRAM-ID-RAUTUSN.
 AUTHCR-MORTON D.
 REMARKS-THIS PROGRAM IS PART OF MY MASTER OF SCIENCE IN SAMPLE
 MANAGEMENT THESIS AND IS DESIGNED TO IDENTIFY SHORT SAMPLE
 TEXTSEGMENT OF A LANGUAGE AS ENGLISH OR SPANISH. THE
 DECISION IS BASED ON STATISTICAL CHARACTERISTICS OF LETTERS
 WITHIN INDEPENDENT WORDS WITHIN CONSIDERATION GIVEN TO
 SPACES.
 INPUT DATA MUST APPEAR IN THE FOLLOWING ORDER:
 3 ENGLISH A CARDS - ICRF TABLE VALUES X(2)/X(1)
 99 ENGLISH B CARDS - DCRF TABLE VALUES X(3)/X(1)
 3 ENGLISH C CARDS - DCRF TABLE VALUES X(2)/X(1)
 99 SPANISH A CARDS - ICRF TABLE VALUES X(2)/X(1)
 99 SPANISH B CARDS - DCRF TABLE VALUES X(3)/X(1)
 99 SPANISH C CARDS - DCRF TABLE VALUES X(2)/X(1)
 1 HEAD CARD PRECEDING EACH GROUP OF SAMPLE CARDS.
 MAXIMUM NUMBER OF SAMPLE CARDS PER TEST IS 9,999 OR 9,99
 INDIVIDUAL CHARACTERS, NOT COUNTING SPACES.
 ENVIRONMENT DIVISION.
 CONFIGURATION SECTION.
 IBM-360-67.
 SUBJECT-COMPUTER.
 SUBJECT-COMPUTER.
 SPECIFIC-INPUT SECTION.
 IBM-360-67.
 INPUT-OUTPUT SECTION.
 FILE-SELECT SAMPLE ASSIGN TO UR-S-INIT.
 SELECT RESULT ASSIGN TO UR-S-OUT.
 DATA SECTION.
 FD SAMPLE RECORDS ARE OMITTED
 BULK RECORDS 5 RECORDS
 DATA RECORD IS SAMPLE-CARD, HEAD-CARD, DATA-CARD.
 01 SAMPLE FILLER.
 05 FILLER-FLD.
 05 FILLER-CB.
 05 FILLER-FR.
 01 HEAD-CARD.
 05 CARD-TYPE PIC X.
 01 NEXT-SAMPLE VALUE PIC X.
 05 FILLER

The image displays a continuous, repeating pattern of binary digits (0s and 1s) arranged in a grid-like structure. The pattern consists of several distinct elements: a top row of '0's, followed by a row of '1's, then a series of '0's and '1's in a repeating sequence. Below this, there are rows of '0's and '1's that form a repeating sequence of symbols, possibly representing a specific data format or a unique identifier. The entire pattern is set against a white background.

016650	01	05 FILLER.	PIC X(18) VALUE ALL ' '.
00167500	01	05 F-LINE-FILLER.	PIC X(26) VALUE SPACES.
00168000	05	FILLER	PIC X(24) VALUE SPACES.
00163500	05	FILLER	PIC X(18) VALUE IDENTIFIED BY YULE'.
00169500-	05	FILLER AS::.	PIC X(30) VALUE QUOTE IS K FOR SINGLE LETTER.
00170500-	05	BANK-ONE	PIC X(30) VALUE SPACES.
00171500	01	05 F-LINE-TWO.	PIC X(30) VALUE SPACES.
00172500	05	FILLER	PIC X(24) VALUE SPACES.
00173500	05	FILLER	PIC X(18) VALUE IDENTIFIED BY YULE'.
00173500-	05	FILLER	PIC X(30) VALUE QUOTE IS K FOR X(2)/X(1) AS:
00174500-	05	BANK-TWO	PIC X(30) VALUE SPACES.
00175500	01	05 FILLER	PIC X(24) VALUE SPACES.
00176500-	05	FILLER	PIC X(18) VALUE IDENTIFIED BY YULE'.
00177500-	05	FILLER	PIC X(30) VALUE QUOTE IS K FOR X(3)/X(1) AS:
00178500-	05	BANK-THREE	PIC X(30) VALUE SPACES.
00179500-	05	FILLER	PIC X(24) VALUE SPACES.
00180500-	05	FILLER	PIC X(49) VALUE SPACES.
00181500-	05	ST FOR VOW-CONS ORDER AS::.	PIC X(30) VALUE SPACES.
00182500-	05	BANK-FOUR	PIC X(30) VALUE SPACES.
00183500-	05	FILLER	PIC X(24) VALUE SPACES.
00184500-	05	ST FOR SINGLETTERS AS::.	PIC X(49) VALUE SPACES.
00185500-	05	BANK-FIVE	PIC X(30) VALUE SPACES.
00186500-	05	FILLER	PIC X(24) VALUE SPACES.
00187500-	05	LINE-SIX	PIC X(49) VALUE SPACES.
00188500-	05	FILLER	PIC X(24) VALUE SPACES.
00189500-	05	BANK-SIX	PIC X(30) VALUE SPACES.
00189500-	05	FILLER	PIC X(24) VALUE SPACES.
00190000-	05	LINE-SEVEN.	PIC X(49) VALUE SPACES.
	01	05 FILLER	PIC X(30) VALUE SPACES.
	01	05 FILLER	PIC X(24) VALUE SPACES.
	01	05 FILLER	PIC X(49) VALUE SPACES.

021450 PERFORM SPAN-FILL-3 1089 TIMES.
 MOVE ZEROS TO CARD-NR.
 BEGIN READ SAMPLE AT END GO TO ECJ CLOSE-CUT-RT THRU INITIALIZE-RT.
 PERFORM DATA-SHIFT VARYING X FROM 1 BY 1 UNTIL X > 71.
 ADD 1 TO CARD-NR ON SIZE ERROR GO TO ALERT.
 MOVE 3 TO N.
 DATA-CCLL (N) = SPACE GO TO BLANK-COL-STEP.
 021950 MOVE 1 TO A.
 CK-CNE IF CC (N) = LTR (A) GO TO PROC-TWO.
 IF A = 33 GO TO COL-STEP.
 ADD 1 TO A.
 GO TO CK-CNE.
 PROC-TWO ADD 1 TO ILF (A) ON SIZE ERROR GO TO ALERT.
 COMPUTE M = N + 1.
 IF CC (M) = SPACE GO TO COL-STEP.
 IF CC (M) = * AND M = 73 PERFORM CARD-STEP-ONE.
 MOVE 1 TO B.
 CK-TWO IF CC (M) = LTR (6) GO TO PROC-THREE.
 IF B = 33 GO TO COL-STEP.
 ADD 1 TO B.
 GO TO CK-TWO.
 PROC-THREE ADD 1 TO FOL-FREQ (A, B) ON SIZE ERROR GO TO ALERT.
 IF A < 11 AND B < 11 ADD 1 TO F-VOW-VCW ON SIZE
 IF A ERROR GO TO ALERT.
 IF A < 11 AND B NOT < 11 ADD 1 TO F-CONS-VOW ON SIZE
 IF A NOT < 11 AND B < 11 ADD 1 TO F-VOW-CCNS ON SIZE
 IF A NOT < 11 AND B NOT < 11 ADD 1 TO F-CONS-CCNS ON SIZE
 IF A ERROR GO TO ALERT.
 COMPUTE L = N + 2.
 IF CC (L) = SPACE GO TO COL-STEP.
 IF CC (L) = * AND L = 73 PERFORM CARD-STEP-TWO.
 MOVE 1 TO B.
 CK-THREE MOVE 1 TO B.
 0223400 IF CC (L) = LTR (B) PERFORM ADDRESS
 0223450 IF CC (L) = LTR (B) GO TO COL-STEP.
 IF B = 33 GO TO B.
 ADD 1 TO B.
 0223750 GO TO CK-THREE.
 0223800 COL-STEP.

```

023850 IF N NOT < 72 GO TO BEGIN.
023950 ADD 1 TO DATA-COLL.
024000 ECJ. PERFORM CLOSE-OUT-RT.
024100 MOVE SPACES TO PRINT-LINE AFTER PAGE-TOP.
024150 WRITE SAMPLE-LINE, RESULT.
024200 CLOSE RUN-STEP.
024250 STOP RUN-STEP.
024300 BLANK-COL-STEP.
024350 ADD 1 TO N.
024400 IF N = 73 GO TO BEGIN.
024450 GO TO DATA-COLL.
024500 CARD-STEP-DONE.
024550 MOVE CC (72) TO CC (2).
024600 READ SAMPLE AT END GO TO ECJ.
024650 PERFORM DAT-SHIFT VARYING X FROM 1 BY 1 UNTIL X > 71.
024700 ADD 1 TO CARD-NR ON SIZE ERROR GO TO ALERT.
024750 COMPUTE N = N + 1.
024800 CARD-STEP-TWO.
024850 MOVE CC (71) TO CC (1).
024900 MOVE CC (72) TO CC (2).
024950 READ SAMPLE AT END GO TO ECJ.
025000 PERFORM DAT-SHIFT VARYING X FROM 1 BY 1 UNTIL X > 71.
025050 ADD 1 TO CARD-NR ON SIZE ERROR GO TO ALERT.
025100 COMPUTE L = N + 2.
025150 ADER COMPUTE B = B + 33.
025200 ADD 1 TO FOL-FREQ (A, B) ON SIZE ERROR GO TO ALERT.
025250 COMPUTE B = B - 33.
025300 DAT-SHIFT.
025350 COMPUTE Y = X + 2.
025400 MOVE CB (X) TO CC (Y).
025450 ENG-FILL-1.
025500 IF L = 0 READ SAMPLE AT END GO TO DEFAULT.
025550 IF L = 0 ADD 1 TO CARD-NR PERFORM ENG-CK-1.
025600 ADD 1 TO L.
025650 MOVE CF-IN (L) TO ENG-ICRF (A).
025700 ADD 1 TO A.
025750 IF L = 11 MOVE ZERO TO L.
025800 ENG-CK-1.
025850 IF DC-ID NOT = 'A' OR DC-SEQ NOT = CARC-NR
025900 OR DC-LANG NOT = ENGLISH P. GO TO DEFAULT.
026000 Eng-FILL-2.
026050 IF L = 0 READ SAMPLE AT END GO TO DEFAULT.
026100 IF L = 0 ADD 1 TO CARD-NR PERFORM ENG-CK-2.
026150
026200

```

```

026250
026300
026400
026450
026500
026550
026600
026650
026700
026750
026800
026850
026900
026950
027000
027100
027150
027200
027250
027300
027350
027400
027450
027500
027550
027600
027650
027700
027750
027800
027850
027900
027950
028000
028050
028100
028150
028200
028250
028300
028350
028400
028450
028500
028550
028600

ADD 1 TO LN (L) TO ENG-DCRF (A, B).
IF B = 33,
  MOVE CRF-IN (L) TO ENG-DCRF (A, B).
  ADD 1 TO A.
  MOVE ZERO TO B.
  ADD 1 TO B MOVE ZERO TO L.
  IF DC-ID NOT = 'B' OR DC-SEQ NOT = CARD-NR
    OR DC-LANG NOT = SPANISH P. GO TO DEFAULT.
  ENG-FILL-3.
  IF L = 0 READ SAMPLE AT END GO TO DEFAULT.
  IF L = 0 ADD 1 TO CARD-NR PERFORM ENG-CK-3.
  ADD 1 TO L
  MOVE CRF-IN (L) TO ENG-DCRF (A, B).
  IF B = 66
    ADD 1 TO A.
    MOVE 33 TO A.
    ADD 1 TO B.
    MOVE ZERO TO L.
  ENG-CK-3.
  IF DC-ID NOT = 'C' OR DC-SEQ NOT = CARD-NR
    OR DC-LANG NOT = SPANISH P. GO TO DEFAULT.
  SPAN-FILL-1.
  IF L = 0 READ SAMPLE AT END GO TO DEFAULT.
  IF L = 0 ADD 1 TO CARD-NR PERFORM SPAN-CK-1.
  ADD 1 TO L
  MOVE CRF-IN (L) TO SPAN-ICRF (A).
  ADD 1 TO A.
  IF L = 11 MOVE ZERO TO L.
  SPAN-CK-1.
  IF DC-ID NOT = 'A' OR DC-SEQ NOT = CARD-NR
    OR DC-LANG NOT = SPANISH P. GO TO DEFAULT.
  SPAN-FILL-2.
  IF L = 0 READ SAMPLE AT END GO TO DEFAULT.
  IF L = 0 ADD 1 TO CARD-NR PERFORM SPAN-CK-2.
  ADD 1 TO L
  MOVE CRF-IN (L) TO SPAN-DCRF (A, B).
  IF B = 33,
    ADD 1 TO A.
    MOVE ZERO TO B.
    ADD 1 TO B.
    MOVE ZERO TO L.
  SPAN-CK-2.
  IF DC-ID NOT = 'B' OR DC-SEQ NOT = SPANISH P. GO TO DEFAULT.
  SPAN-FILL-3.
  IF L = 0 READ SAMPLE AT END GO TO DEFAULT.

```

```

028650 IF L = 0 ADD 1 TO CARD-NR PERFORM SPAN-CK-3.
028700 ADD 1 TO LN (L) TO SPAN-DCRF (A, B).
028800 IF B = 66,
028850 ADD 1 TO A.
028900 MOVE 33 TO B.
028950 ADD 1 TO B MOVE ZERO TO L.
029000 SPAN-CK-3.
029050 IF DC-ID NOT = 'C' OR DC-SEQ-NCT = CARD-NR
029100 SR DC-LANG NCT = SPANISH P. GO TO DEFAULT.
029150 MOVE ! RUN TERMINATED DUE TO INPUT SEQUENCE ERROR • TC
029200 PRINT-LINE.
029250 WRITE PRINT-LINE AFTER PAGE-TOP.
029300 CLOSE SAMPLE, RESULT.
029350 STOP RUN.

ALTER* WRITE PRINT-LINE FROM ERROR-LINE AFTER PAGE-TOP.
CYCLE* READ SAMPLE AT END GO TO ECJ.
IF NEXT-SAMPLE GO TO RESART.
GO TO CYCLE.

FESTART* PERFORM INITIALIZER-T
PERFORM DATA-SHIFT VARYING X FROM 1 BY 1 UNTIL X > 71.
ADD 1 TO CARD-NR.
MOVE 3 TO CAN.
GO TO DATA-COLL.

CLOSE OUT-RT-NR = 0 GO TO INITIALIZER-RT
PERFORM IND-TOTAL-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
PERFORM IND-PROB-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
PERFORM IND-PREP-CALC VARYING A FROM 1 BY 1 UNTIL A > 33.
COMPUTE K-ROUND = (10000 * (S - TLT-TOTAL)) /
COMPUTE T-TOTAL = ((K - 645.079) * (CARD-NR ** .5)) /
COMPUTE Z-ROUND = ((K - 751.447) * (CARD-NR ** .5)) /
PERFORM INI-LANG-CRK-EXIT
PERFORM PRINT-LINE FROM TITLE-UL BEFORE PAGE-TOP.
WRITE PRINT-LINE FROM COL-HEAD BEFORE Q.
WRITE PRINT-LINE FROM COL-HEAD BEFORE 2.
MOVE BANK-FDRY TO BANK-CNE.
WRITE PRINT-LINE FROM C-LINE-CNE BEFORE 2.
PERFORM CRILP-CALC VARYING A FROM 2 BY 1 UNTIL A > 33.

```

```

031050 PERFORM DEPN-TWO-SUM VARYING A FROM 1 BY 1 UNTIL A > 33
031100 AFTER B FROM 1 BY 1 UNTIL B > 33 * 1 UNTIL A > 33
031150 PERFORM DEPN-TWO-PROB VARYING A FROM 1 BY 1 UNTIL A > 33
031200 AFTER B FROM 1 BY 1 UNTIL B > 33.

MOVE ZEROS TO TOS-K-PREP VARYING A FROM 1 BY 1 UNTIL A > 33
AFTER B FROM 1 BY 1 UNTIL B > 33 * 1 UNTIL A > 33
COMPUTE K ROUNDED = (10000 * (S - FOL-FREQ-TOTAL)) /
(FOL-FREQ-TOTAL ** 2) *
COMPUTE Z-FENG ROUNDED = (K - 87.168) * (CARD-NR ** .5) /
40.207 *
COMPUTE Z-SPAN ROUNDED = (K - 104.553) * (CARD-NR ** .5) /
38.094 *
PERFORM LANG-CK-K THRU LANG-CK-Y-EXIT.

MOVE BANK-FORM TO BANK-TWO
WHITE PRINT-LINE FROM O-LINE-TWO BEFORE 2* BY 1 UNTIL A > 33
PERFORM CRT-FOL-PROG-CALC VARYING A FROM 1 BY 1 UNTIL B > 33.
AFTER 3 FROM 1 BY 1 UNTIL B > 33.

MOVE FOL-FREQ-TOTAL-FREQ-TOTAL * CRT-FOL-FREQ-TOTAL * CRT-FOL-FREQ-TOTAL
MOVE ZEROES DEPN-TWO-SUM VARYING A FROM 1 BY 1 UNTIL A > 33
AFTER B FROM 34 BY 1 UNTIL B > 66 *
PERFORM DEPN-TWO-PROB VARYING A FROM 1 BY 1 UNTIL A > 33
AFTER B FROM 34 BY 1 UNTIL B > 66.

MOVE ZEROS TOS-K-PREP VARYING A FROM 1 BY 1 UNTIL A > 33
AFTER B FROM 34 BY 1 UNTIL B > 66 *
COMPUTE K ROUNDED = (10000 * (S - FUL-FREQ-TOTAL)) /
(FUL-FREQ-TOTAL ** 2) *
COMPUTE Z-FENG ROUNDED = (K - 62.083) * (CARD-NR ** .5) /
42.821 *
COMPUTE Z-SPAN ROUNDED = (K - 72.68) * (CARD-NR ** .5) /
35.210 *
PERFORM LANG-CK-K THRU LANG-CK-K-EXIT.

MOVE BANK-FORM TO BANK-THREE
WHITE PRINT-LINE FROM C-LINE-THREE BEFORE 2* BY 1 UNTIL A > 33
PERFORM CRT-FOL-PROG-CALC VARYING A FROM 1 BY 1 UNTIL B > 66 *
COMPUTE VC-TOTAL = F-CONS-CCNS + F-VOW-CONS + F-CONS-VOW
+ F-VOW-VOW ON SIZE ERROR GO TO ALERT.
IF F-VOW-VOW > 0 COMPUTE VC-CRF (1) ROUNDED = F-VOW-VOW / VC-TOTAL
ELSE MOVE ZEROS TO VC-CRF (1).
IF F-CONS-VOW > 0 COMPUTE VC-CRF (2) ROUNDED = (F-CONS-VOW / VC-TOTAL)
+ VC-CRF (1)
ELSE MOVE VC-CRF (1) TO VC-CRF (2).
IF F-VOW-CONS > 0

```


PERFORM ZEROING-TWO VARYING A FROM 1 BY 1 UNTIL A > 33
AFTER B FROM 1 BY 1 UNTIL B > 66.


```

040650
040700 YULES-K-PREP = S + (FOL-FREQ (A, B) ** 2).
040800 TO FOL-PRCB (A, B).
040850 CR-CMPUTER-CALC = CRF-TOTAL + FOL-PRCB (A, B).
040900 MOVE CRF-TOTAL TO FOL-PROB (A, B).
040950 ENG-VC-DCK.
041000 COMPUTE D = ENG-VC-CRF (A) - VC-CRF (A).
041050 IF D > ENG-D MOVE D TO ENG-D.
041100 SPAN-CMPUTE D = SPAN-VC-CRF (A) - VC-CRF (A).
041150 IF D > SPAN-D MOVE D TO SPAN-D.
041200 LANG-CK-KS.
041250 IF ENG-D = SPAN-D * NO DEC SN* TO BANK-LANG,
041300 MOVE ENG-D TO D.
041350 IF ENG-D > SPAN-D! TO BANK-LANG,
041400 MOVE SPAN-D TO D.
041450 IF ENG-D > SPAN-D! TO BANK-LANG,
041500 MOVE SPAN-D TO D.
041550 COMPUTE ENG-D TO D.
041600 COMPUTE Z2 ROUNDED = SAMPLE-SIZE ** .5.
041650 COMPUTE E ROUNDED = D * Z2.
041700 IF E < 2.8? * MOVE KS-ALFA (1) TO BANK-EQUAL.
041750 IF E > 1.82? TO BANK-EQUAL' ALFA.
042000 MOVE KS-ALFA (155) TO BANK-ALFA.
042100 IF E > 1.9 AND E < 1.93, COMPUTE EOUNDED = (100 * E) - 27,
042150 MOVE KS-ALFA (1) TO BANK-ALFA,
042200 MOVE SPACES TO BANK-EQUAL.
042250 ENG-KS-CK1.
042300 COMPUTE D = ENG-ICRF (A) - ILP (A).
042350 IF D > ENG-D MOVE D TO ENG-D.
042400 SPAN-KS-CK1.
042450 COMPUTE D = SPAN-ICRF (A) - ILP (A).
042500 IF D > SPAN-D MOVE D TO SPAN-D.
042550 ENG-KS-CK2.
042600 COMPUTE D = ENG-DCRF (A, B) - FOL-PROB (A, B).
042650 IF D > ENG-D MOVE D TO ENG-D.
042700 SPAN-KS-CK2.
042750 COMPUTE D = SPAN-DCRF (A, B) - FOL-PROB (A, B).
042800 IF D > SPAN-D MOVE D TO SPAN-D.
042850 ZERGING-ONE.
043000

```

IDEN-ONE
IDEN-ONE
IDEN-ONE
IDEN-ONE

043050 MOVE ZEROS TO ILF (A), ILP (A).
043100 ZEROING-TWO MOVE ZEROS TO FOL-FREQ (A, E), FOL-PROB (A, B).

APPENDIX J. LANGUAGE IDENTIFICATION PROGRAM (V-2)

IDENTIFICATION DIVISION.
 PROGRAM-ID. LANGUAGE-VERSION-TWO.
 AUTHOR. MORTON D. RAU, USN.
 REMARKS. THIS PROGRAM IS PART OF MY MASTER OF SCIENCE IN SAMPLE
 TEXT OF A LANGUAGE AS DESIGNED TO IDENTIFY SHORT THE
 DECISION IS BASED ON STATISTICAL CHARACTERISTICS OF LETTERS,
 THE SPACE AND THE INTERDEPENDENCY OF LETTERS.
 ADJUSTING WORDS MUST APPEAR IN THE FOLLOWING ORDER:
 4 ENGLISH CARDS - DCRF TABLE VALUES X(2)/X(1)
 106 ENGLISH CARDS - DCRF TABLE VALUES X(3)/X(1)
 4 SPANISH CARDS - DCRF TABLE VALUES X(2)/X(1)
 106 SPANISH CARDS - DCRF TABLE VALUES X(3)/X(1)
 106 SPANISH CARD PRECEDED IN EACH GROUP OF SAMPLE CARDS.
 MAXIMUM NUMBER OF SAMPLE CARDS PER TEST IS 999 OR 9,999.
 INDIVIDUAL CHARACTERS, INCLUDING SPACES.
 ENVIRONMENT SECTION.
 CONFIGURATION COMPUTER. IBM-360-67.
 SOURCE-COMPUTER. IBM-360-67.
 SPECIAL-COMPUTER.
 INPUT-OUTPUT SECTION.
 FILE CONTROL SECTION.
 SELECT SAMPLE ASSIGN TO UR-S-IN1.
 SPECIFIC SECTION.
 INPUT-OUTPUT SECTION.
 FILE SECTION.
 DATA SECTION.
 FD SAMPLE RECORDS ARE OMITTED
 LABELED RECORDS IS SAMPLE-CARD, HEAD-CARD, DATA-CARD.
 01 DATA-CARD.
 05 FILLER.
 05 SAMPLE-FLD.
 10 FILLER.
 05 HEAD-CARD.
 05 CARD-TYPE PIC A.
 05 NEXT-SAMPLE VALUE PIC X.
 05 FILLER.
 01 HEAD-CARD.
 05 CARD-TYPE PIC A.
 05 NEXT-SAMPLE VALUE PIC X.
 05 FILLER.

VALUE

କରିବାକୁ ପାଇଁ ଏହାର ମଧ୍ୟ ଦେଖିଲୁ କିମ୍ବା କିମ୍ବା କିମ୍ବା କିମ୍ବା କିମ୍ବା

oooooooooooooooooooooooooooooooooooo

如若說這就是我們的命運，那我們就只能說：我們的命運是悲慘的。

016650	05	FILLER	PIC X(7) VALUE SPACES.
016700	05	FILLER	PIC X(18) VALUE ALL ''.
016750	05	0-LINETHREE.	PIC X(26) VALUE SPACES.
016800	05	FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY YULE'.
016850	05	FILLER	PIC X(18) VALUE QUOTE 'S K FOR SINGLE LETTER
016900	05	FILLER	PIC X(30) VALUE QUOTE 'S K FOR SINGLE LETTER
016950	05	FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY YULE'.
017000	05	S BANK-ONE	PIC X(18) VALUE QUOTE 'S K FOR X(2)/X(1) AS IDENTIFIED BY YULE'.
017050	05	FILLER	PIC X(30) VALUE QUOTE 'S K FOR X(2)/X(1) AS IDENTIFIED BY YULE'.
017100	05	FILLER	PIC X(30) VALUE SPACES.
017150	05	FILLER	PIC X(30) VALUE SPACES.
017200	05	LINETWO.	PIC X(24) VALUE SPACES IDENTIFIED BY YULE'.
017250	05	FILLER	PIC X(18) VALUE QUOTE 'S K FOR X(3)/X(1) AS IDENTIFIED BY YULE'.
017300	05	FILLER	PIC X(30) VALUE QUOTE 'S K FOR X(3)/X(1) AS IDENTIFIED BY YULE'.
017350	05	FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY YULE'.
017400	05	FILLER	PIC X(18) VALUE QUOTE 'S K FOR X(3)/X(1) AS IDENTIFIED BY YULE'.
017450	05	FILLER	PIC X(30) VALUE SPACES.
017500	05	BANK-TWO	PIC X(30) VALUE SPACES.
017550	05	FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY YULE'.
017600	05	0-LINETHREE.	PIC X(49) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
017650	05	FILLER	PIC X(30) VALUE SPACES.
017700	05	FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
017750	05	FILLER	PIC X(49) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
017800	05	FILLER	PIC X(30) VALUE SPACES.
017850	05	BANK-THREE	PIC X(24) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
017900	05	0-LINE-FOUR.	PIC X(49) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018000	05	FILLER	PIC X(30) VALUE SPACES.
018050	05	FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018100	05	0-ST FILLER	PIC X(49) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018150	05	0-ST FILLER	PIC X(30) VALUE SPACES.
018200	05	0-ST FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018250	05	0-ST FILLER	PIC X(49) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018300	05	0-ST FILLER	PIC X(30) VALUE SPACES.
018350	05	0-ST FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018400	05	0-ST FILLER	PIC X(49) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018450	05	0-ST FILLER	PIC X(30) VALUE SPACES.
018500	05	0-ST FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018550	05	0-ST FILLER	PIC X(49) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018600	05	0-ST FILLER	PIC X(30) VALUE SPACES.
018650	05	0-ST FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018700	05	0-ST FILLER	PIC X(49) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018750	05	0-ST FILLER	PIC X(30) VALUE SPACES.
018800	05	05-BANK-SIX	PIC X(24) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018850	05	FILLER	PIC X(49) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
018900	05	05-BANK-SEVEN.	PIC X(30) VALUE SPACES.
018950	05	FILLER	PIC X(24) VALUE SPACES IDENTIFIED BY K-S TEIDEN.
019000	05	05-BANK-EIGHT.	PIC X(49) VALUE SPACES IDENTIFIED BY K-S TEIDEN.


```

023850 IF MATCH = ZERO GO TO CK-THREE.
023900 DELTA COMPUTE B = B + 34
024000 ADD 1 TO FOL-FREQ (A, B) ON SIZE ERROR GO TO ALERT.
024050 ECHO ADD 1 TO N.
024100 CK-DNE ADD 1 TO N.
024150 CK-DNE ADD 1 TO ILF (34) ON SIZE ERROR GO TO ALERT.
024200 ADD 1 TO N.
024250 MOVE 34 TO A.
024300 CK-DNE-ALPHA.
024350 IF CL (N) = 73 PERFORM SHIFT-ONE.
024400 COMPUTE N = N - 1
024450 PERFORM SECOND-LETTER-CHECK THRU SECOND-EXIT.
024500 IF MATCH = 1 GO TO CHARLIE, ELSE ADD 1 TO N.
024550 CK-DNE-BRAVO.
024600 ADD 1 TO N.
024650 CK-TWO ADD 1 TO CK-DNE-ALPHA.
024700 CK-TWO ADD 1 TO FOL-FREQ (A, 34) ON SIZE ERROR GO TO ALERT.
024750 ADD 1 TO N.
024800 CK-TWO-ALPHA.
024850 COMPUTE M = N + 1.
024900 IF N = 73 PERFORM SHIFT-TWO.
024950 IF CL (M) = SPACE GO TO CK-TWO-BRAVO.
025000 COMPUTE N = N - 1
025050 PERFORM THIRD-LETTER-CHECK THRU THIRD-EXIT.
025100 IF MATCH = 1 GO TO DELTA, ELSE ADD 1 TO N.
025150 CK-TWO-BRAVO.
025200 ADD 1 TO N.
025250 CK-TWO-ALPHA.
025300 CK-TWO ADD 1 TO CK-TWO-ALPHA.
025350 GO TO CK-TWO-ALPHA.
025400 SHIFT-ONE.
025450 ADD 1 TO FOL-FREQ (A, 68) ON SIZE ERROR GO TO ALERT.
025500 SHIFT-ONE.
025550 MOVE SPACES TO CL (1), CL (2).
025600 MOVE 3 TO N.
025650 SHIFT-TWO.
025700 MOVE CL (72) TO CL (2).
025750 PERFORM READ-A-CARD.
025800 MOVE 2 TO N.
025850 MOVE 3 TO M.
025900 SHIFT-THREE.
025950 MOVE CL (72) TO CL (2).
026000 PERFORM READ-A-CARD.

```

```

026250 MOVE 1 TO L.
026300 MOVE 3 TO L.
ALT-THREE CL (72) TO CL (1).
0264C9 MOVE SPACES TO CL (2).
PERFORM READ-A-CARD.
026450 MOVE 0 TO N.
026500 MOVE 2 TO L.
026600 ENG-FILL-1 = 0 READ SAMPLE AT END GO TO DEFAULT.
026650 IF L = 0 ADD 1 TO CARD-NR PERFORM ENG-CK-1.
026700 ADD 1 TO L.
026750 MOVE CRF-IN (L) TO ENG-ICRF (A).
026800 ADD 1 TO A.
026850 IF L = 11 MOVE ZERO TO L.
ENG-CK-1 DC-ID NOT = *D* OR DC-SEQ NOT = CARD-NR
026900 OR DC-LANG NOT = •ENGLISH P • GO TO DEFAULT.
026950 ENG-FILL-2 = 0 READ SAMPLE AT END GO TO DEFAULT.
027000 IF L = 0 ADD 1 TO CARD-NR PERFORM ENG-CK-2.
027050 ADD 1 TO L.
027100 MOVE CRF-IN (L) TO ENG-DCRF (A, B).
027150 IF B = 34 ADD 1 TO A.
027200 MOVE ZERO TO B.
027250 ADD 1 TO B.
027300 ENG-FILL-3 = 0 READ SAMPLE AT END GO TO DEFAULT.
027350 IF L = 0 ADD 1 TO CARD-NR PERFORM ENG-CK-3.
027400 MOVE CRF-IN (L) TO ENG-DCRF (A, B).
027450 IF B = 68 ADD 1 TO A.
027500 MOVE 34 TO B.
027550 ADD 1 TO B.
027600 ENG-FILL-4 = 0 READ SAMPLE AT END GO TO DEFAULT.
027650 IF L = 0 ADD 1 TO CARD-NR PERFORM ENG-CK-4.
027700 MOVE CRF-IN (L) TO ENG-DCRF (A, B).
027750 IF B = 68 ADD 1 TO A.
027800 MOVE 34 TO B.
027850 ADD 1 TO B.
027900 ENG-FILL-5 = 0 READ SAMPLE AT END GO TO DEFAULT.
027950 IF L = 0 ADD 1 TO CARD-NR PERFORM ENG-CK-5.
028000 SPAN-FILL-1 = 0 READ SAMPLE AT END GO TO DEFAULT.
028050 IF L = 0 ADD 1 TO CARD-NR PERFORM SPAN-CK-1.
028100 ADD 1 TO L.
028150
028200
028250
028300
028350
028400
028450
028500
028550
028600

```

028650 MOVE CRF-IN (L) TO SPAN-ICRF (A).
 028700 ADD 1 TO A.
 028750 IF L = 11 MOVE ZERO TO L.
 028800 SPAN-CK-1.
 028850 IF DC-ID NOT = 'D' OR DC-SEQ NOT = CARE-NR
 CDR DC-LANG NOT = SPANISH P. GO TO DEFAULT.
 028900 SPAN-FILL-2.
 028950 IF L = 0 READ SAMPLE AT END GO TO DEFAULT.
 029000 ADD 1 TO L.
 029050 ADD 1 TO L.
 029100 MOVE CRF-IN (L) TO SPAN-DCRF (A, B).
 029150 IF B = 34, ADD 1 TO A.
 029200 MOVE ZERO TO B.
 029250 ADD 1 TO B.
 029300 SPAN-CK-2.
 029350 IF DC-ID NOT = 'E' OR DC-SEQ NOT = CARE-NR
 SPAN-FILL-3.
 029400 IF L = C READ SAMPLE AT END GO TO DEFAULT.
 029450 IF L = O ADD 1 TO C CARD-NR PERFORM SPAN-CK-3.
 029500 ADD 1 TO L.
 029550 MOVE CRF-IN (L) TO SPAN-DCRF (A, B).
 029600 IF B = 68, ADD 1 TO A.
 029650 MOVE 34 TO B.
 029700 ADD 1 TO B.
 029750 SPAN-CK-3.
 029800 IF DC-ID NOT = 'F' OR DC-SEQ NOT = CARE-NR
 CDR DC-LANG NOT = SPANISH P. GO TO DEFAULT.
 029850 MOVE * RUN TERMINATED DUE TO INPUT SEQUENCE ERROR * TO
 PRINT-LINE.
 029900 WRITE PRINT-LINE AFTER PAGE-TOP.
 029950 CLOSE SAMPLE, RESULT.
 030000 STOP RUN.
 030050 ALERT * WRITE PRINT-LINE FROM ERROR-LINE AFTER PAGE-TOP.
 030100 CYCLE.
 030150 READ SAMPLE AT END GO TO ECJ.
 030200 IF NEXT-SAMPLE GO TO RESTART.
 030250 GO TO CYCLE.
 030300 RESTART.
 030350 PERFORM INITIALIZE-RT.
 030400 PERFORM DAT-SHIFT VARYING X FROM 1 BY 1 UNTIL X > 71.
 030450 ADD 1 TO CARD-NR.
 030500

```

031050 MOVE 2 TO N.
031150 GO TO BRAVJ.
031250 CLOSE-CUT-BRT.
031350 COMPUTE IND-TOTAL-CALC VARYING A FROM 1 BY 1 UNTIL A > 34.
031450 COMPUTE YULE-PREP VARYING A FROM 1 BY 1 UNTIL A > 34.
031450 COMPUTE K-ROUNDED = (10000 * (S - ILF-TOTAL)) /
031550 COMPUTE Z-ENG-ROUNDED = ((K - 973.97) * (CARD-NR ** .5)) /
031650 COMPUTE Z-SPAN ROUNDED = (K - 806.481) * (CARD-NR ** .5) /
031750 PERFORM PRINT-LINE LANG-CK-K-EXITE-TCP.
031750 WRITE BANK-FROM C-LINE-EONE BEFORE TITLE-UL AFTER PAGE-TCP.
031850 WRITE PRINT-LINE FROM COL-HEAD BEFORE Q.
031950 MOVE BANK-FROM C-LINE-EONE BEFORE TITLE-UL AFTER PAGE-TCP.
032050 WRITE PRINT-LINE CR-LP-CALC VARYING A FROM 2 BY 1 UNTIL A > 34.
032100 PERFORM DEPN-TWO-SUM VARYING A FROM 1 BY 1 UNTIL B > 34.
032150 AFTER BN-FROM LP-BY 1 UNTIL B > 34.
032200 PERFORM DEPN-TWO-PREP VARYING A FROM 1 BY 1 UNTIL B > 34.
032250 MOVE ZEROS TO S-PREP VARYING A FROM 1 BY 1 UNTIL B > 34.
032300 PERFORM YULES-TOTAL-CALC VARYING A FROM 1 BY 1 UNTIL B > 34.
032350 AFTER BN-FROM 1 BY 1 UNTIL B > 34.
032400 COMPUTE K-ROUNDED = (10000 * (S - FOL-FREQ-TOTAL)) /
032450 COMPUTE Z-ENG-ROUNDED = (K - 87.93) * (CARD-NR ** .5) /
032500 COMPUTE Z-SPAN ROUNDED = (K - 108.764) * (CARD-NR ** .5) /
032550 PERFORM LANG-CK-K-EXITE-TCP.
032600 MOVE BANK-FROM C-LINE-EONE BEFORE TITLE-UL AFTER PAGE-TCP.
032650 WRITE PRINT-LINE CR-LP-CALC VARYING A FROM 1 BY 1 UNTIL B > 34.
032700 MOVE FOL-FREQ-TOTAL-CALC VARYING A FROM 1 BY 1 UNTIL B > 34.
032750 AFTER BN-FROM LP-BY 1 UNTIL B > 34.
032800 PERFORM DEPN-TWO-SUM VARYING A FROM 1 BY 1 UNTIL B > 68.
032850 MOVE ZEROS TO S-PREP VARYING A FROM 1 BY 1 UNTIL B > 68.
032900 PERFORM YULES-TOTAL-CALC VARYING A FROM 1 BY 1 UNTIL B > 68.
032950 AFTER BN-FROM LP-BY 1 UNTIL B > 68.
033000 MOVE ZEROS TO S-PREP VARYING A FROM 1 BY 1 UNTIL B > 68.
033050 PERFORM DEPN-TWO-SUM VARYING A FROM 1 BY 1 UNTIL B > 68.
033100 MOVE ZEROS TO S-PREP VARYING A FROM 1 BY 1 UNTIL B > 68.
033150 PERFORM YULES-TOTAL-CALC VARYING A FROM 1 BY 1 UNTIL B > 68.
033200 AFTER BN-FROM LP-BY 1 UNTIL B > 68.
033250 PERFORM DEPN-TWO-SUM VARYING A FROM 1 BY 1 UNTIL B > 68.
033300 MOVE ZEROS TO S-PREP VARYING A FROM 1 BY 1 UNTIL B > 68.
033350 PERFORM YULES-TOTAL-CALC VARYING A FROM 1 BY 1 UNTIL B > 68.

```

```

033450 COMPUTE K ROUNDED = (10000 * (S - FCL-FREQ-TOTAL)) /
033500 (FOL-FREQ-TOTAL *# 2) / IDENT-TWO
033550 COMPUTE Z-ENG ROUNDED = (K - 73.139) * (CARD-NR ** .5) / IDENT-TWO
033600 COMPUTE Z-SPAN ROUNDED = (K - 83.553) * (CARD-NR ** .5) / IDENT-TWO
033700 PERFORM LANG-CK-K THRU LANG-CK-K-EXIT.
033750 MOVE BANK-FORM TO BANK-THREE.
033800 WRITE PRINT-O-LINE-THREE BEFORE 2.
033850 PERFORM CP-FOL-PROB-CALL-VARYING A FROM 1 BY 1 UNTIL A > 34.
033900 AFTER 3 FROM 35 BY 1 UNTIL B > 68.
033950 COMPUTE VC-TOTAL = F-CONS-CCNS + F-VCW-CONS + F-CONS-VOW
034000 + F-VCW-VOW ON SIZE ERROR GO TO ALERT.
034100 IF F-VCW-VOW > 0 COMPUTE VC-CRF (1) ROUNDED = F-VCW-VOW / VC-TOTAL
034150 ELSE MOVE VC-ZEROS TO VC-CRF (1).
034200 IF F-CONS-VC-VOW > 0 COMPUTE VC-CRF (2) ROUNDED = (F-CONS-VOW / VC-TOTAL)
034250 ELSE MOVE VC-CRF (1) TO VC-CRF (2).
034300 IF F-VCW-CONS > 0 COMPUTE VC-CRF (3) ROUNDED = (F-VCW-CONS / VC-TOTAL)
034350 ELSE MOVE VC-CRF (2) TO VC-CRF (3).
034400 IF F-CONS-CCNS > 0 COMPUTE VC-CRF (4) ROUNDED = (F-CONS-CONS / VC-TOTAL)
034450 ELSE MOVE VC-CRF (3) TO VC-CRF (4).
034500 PERFORMANCE ENG-VC-CK VARYING A FROM 1 BY 1 UNTIL A > 4.
034550 MOVE VC-TOTAL TO SAMPLE-SIZE.
034600 MOVE LANG-CK-KS.
034650 MOVE BANK-FDRM TO BANK-FOUR.
034700 WRITE PRINT-O-LINE-FOUR BEFORE 2.
034750 MOVE ZEROES TO SPAN-D.
034800 PERFORMANCE ENG-KS-CK1 VARYING A FROM 1 BY 1 UNTIL A > 34.
034850 PERFORMANCE SPAN-KS-CK1 VARYING A FROM 1 BY 1 UNTIL A > 34.
034900 MOVE TLF-STOTAL TO SAMPLE-SIZE.
034950 MOVE LANG-CK-KS.
035000 MOVE BANK-FDRM TO BANK-FIVE.
035050 WRITE PRINT-O-LINE-FIVE BEFORE 2.
035100 MOVE ZEROES TO SPAN-D.
035150 PERFORMANCE ENG-KS-CK2 VARYING A FROM 1 BY 1 UNTIL A > 34.
035200 AFTER 3 FROM 1 BY 1 UNTIL B > 34.
035250 PERFORMANCE SPAN-KS-CK2 VARYING A FROM 1 BY 1 UNTIL A > 34.
035300 MOVE FFT2 TO SAMPLE-SIZE.
035350 MOVE LANG-CK-KS.

```



```

040650 COMPUTE P ROUNDED = (P / 62.03) * 22.0
040700 COMPUTE J = J + P CN SIZE ERROR GO TO ALERT
040750 COMPUTE BANK-ALFA ROUNDED = I - (2 * (2 * .3989422804014 * 
040800 MOVE SPACES TO BANK-EQUAL.
040850 LANG-K-EXIT.
040900 CRILP-CALC-E B = A - 1. ILP (B) + ILP (A).
040950 COMPUTE ILP (A) = IND-TOTAL-CALC.
041000 IND-CALC-TOTAL = ILF-TOTAL + ILF (A).
041050 IND-PROCB-CALC-I
IF ILF (A) > 0 COMPUTE ILP (A) ROUNDED =
(ILF (A) / ILF-TOTAL), ELSE MOVE ZEROS TO ILP (A).
041100 YULE-PREP S = S + (ILF (A) ** 2).
041150 DEPN-TOTAL-SUM = COMPUTE FOL-FREQ-TOTAL = FOL-FREQ-TOTAL + FOL-FREQ (A, B).
041200 DEPN-TWC-PRCB IF FOL-FREQ (A, B) > 0 COMPUTE FOL-PPCB (A, B) ROUNDED =
(FOL-FREQ (A, B) / FOL-FREQ-TOTAL), ELSE MOVE ZEROS
TO FOL-PROB (A, B).
041250 YULEST-K-PREP S = S + (FOL-FREQ (A, B) ** 2).
041300 CR-FOL-PROB-CALC TOTAL = CRF-TOTAL + FOL-PPCB (A, B).
041350 MOVE CRF-TOTAL TO FOL-PRDB (A, B).
041400 ENG-VC-DOCK COMPUTE D = ENG-VC-CRF (A) - VC-CRF (A).
IF D > ENG-D MOVE D TO ENG-D.
041450 SPAN-VC-DOCK COMPUTE D = SPAN-VC-CRF (A) - VC-CRF (A).
IF D > SPAN-D MOVE D TO SPAN-D.
041500 LANG-KS IF ENG-D = SPAN-D MOVE *NO DEC$N* TO BANK-LANG,
041550 MOVE ENG-D TO D.
IF ENG-D > SPAN-D MOVE *SPANISH* TO BANK-LANG,
041600 MOVE SPAN-D TO D.
IF ENG-D < SPAN-D MOVE *ENGLISH* TO BANK-LANG,
041650 MOVE ENG-D TO D.
COMPUTE Z2 ROUNDED = SAMPLE-SIZE ** .5.
COMPUTE E ROUNDED = D * Z2.
IF E < 26 MOVE *26* TO BANK-EQUAL,
MOVE KS-ALFA (1) TO BANK-ALFA.
041700
041750
041800
041850
041900
041950
042000
042050
042100
042150
042200
042250
042300
042350
042400
042450
042500
042550
042600
042650
042700
042750
042800
042850
042900
042950
043000

```

```

043050 IF E > 1.82 ! TO BANK-EQUAL.
043100 MOVE KS-ALFA (155) TO BANK-ALFA.
043150 IF E > 27 AND E < 1.83, MOVE KS-TALFA (T) TO BANK-ALFA, - 27,
043200 COMPUTE T ROUNDED = (100 * E) -
043250 MOVE SPACES TO BANK-EQUAL.
043300 ENG-KS-CK1 = ENG-ICRF (A) - ILP (A).
043350 SPAN-KS-CK1.
IF D > ENG-D MOVE D TO ENG-D.
043400 COMPUTE D = SPAN-ICRF (A) - ILP (A).
043450 ENG-KS-CK2.
IF D > SPAN-D MOVED TO SPAN-D.
043500 COMPUTE D = ENG-DCRF (A, B) - FOL-PRCB (A, B).
043550 SPAN-KS-CK2.
IF D > ENG-D MOVE D TO ENG-D.
043600 COMPUTE D = SPAN-DCRF (A, B) - FOL-PROB (A, B).
043650 SPAN-KS-CK2.
IF D > SPAN-D MOVED TO SPAN-D.
043700 ZEROING ONE ZERO TO ILF (A), ILP (A).
043750 ZEROING TWO ZEROES TO FOL-FREQ (A, B), FCL-PROB (A, B).
043800 READ-A-CARD READ SAMPLE AT END GO TO ECJ.
043850 IF NEXT-SAMPLE GO TO FLUSH.
043900 ADD 1 TO CARD-NR ON SIZE ERROR GO TO ALERT.
043950 PERFORM DAT-SHIFT VARYING X FROM 1 BY 1 UNTIL X > 71.
044000 LAT-SHIFT X COMPUTE Y = X + 2.
044050 MOVE CC (X) TO CL (Y).
044100 FLUSH.
IF CARD-NR NOT = ZERO PERFORM CLOSE-OUT-RT.
044150 PERFORM INITIALIZ-RT.
GO TO ALPHA.
044200 ECJ.
044250 MOVE SPACES TO PRINT-LINE.
CLOSE SAMPLE, RESULT.
044300 CLOSE RUN.
044350 FIRST-LETTER-CHECK.
MOVE 1 TO A.
044400 MOVE ZERO TO MATCH.
044450 FLC-ALPHA.
IF CL (N) = LTR (A) GO TO FLC-BRAVO.
IF A = 32 GO TO FIRST-EXIT.
ADD 1 TO A.

```

```

045450 GO TO FLC-ALPHA.
045550 MOVE I TO MATCH.
045650 FIRST-EXIT.
045750 SECOND-LETTER-CHECK.
045850 MOVE 1 TO B TO MATCH.
045950 MOVE ZERO TO N + 1.
046050 SLC-ALPHA. IF CL (M) = LTR (B) GO TO SLC-BRAVO.
046150 IF B = 33 GO TO SECOND-EXIT.
046250 ADD 1 TO B.
046350 GO TO SLC-ALPHA.
046450 SLC-BRAVO. MOVE I TO MATCH.
046550 MOVE 1 TO B TO MATCH.
046650 MOVE 1 TO B TO MATCH.
046750 MOVE ZERO TO N + 2.
046850 TLC-ALPHA. IF CL (L) = LTR (B) GO TO TLC-BRAVO.
046950 IF B = 33 GO TO THIRD-EXIT.
046A50 ADD 1 TO B.
046B50 GO TO TLC-ALPHA.
046C50 MOVE I TO MATCH.
046D50 MOVE 1 TO MATCH.
046E50 VC-TALLEY. IF A = 34 GO TO VC-EXIT.
046F50 IF A < 11 AND B < 11 AND C < 11 AND D < 11 AND E < 11 AND F < 11 AND G < 11 AND H < 11 AND I < 11 AND J < 11 AND K < 11 AND L < 11 AND M < 11 AND N < 11 AND O < 11 AND P < 11 AND Q < 11 AND R < 11 AND S < 11 AND T < 11 AND U < 11 AND V < 11 AND W < 11 AND X < 11 AND Y < 11 AND Z < 11 ADD 1 TO F-CONS-VCW ON SIZE
047050 IF A < 11 AND B < 11 AND C < 11 AND D < 11 AND E < 11 AND F < 11 AND G < 11 AND H < 11 AND I < 11 AND J < 11 AND K < 11 AND L < 11 AND M < 11 AND N < 11 AND O < 11 AND P < 11 AND Q < 11 AND R < 11 AND S < 11 AND T < 11 AND U < 11 AND V < 11 AND W < 11 AND X < 11 AND Y < 11 AND Z < 11 ADD 1 TO F-VCW-CONS ON SIZE
047150 IF A NOT < 11 AND B < 11 ADD 1 TO F-VCW-CONS ON SIZE
047250 IF A ERROR GO TO ALERT.
047350 IF A NOT < 11 AND B NOT < 11 ADD 1 TO F-CONS-CONS ON SIZE
047450 VC-EXIT. EXIT.

```

LIST OF REFERENCES

1. Nakamura, Y., "Identification of Languages with Short Sample Texts: A Linguometric Study," Library and Information Science, No. 9, pp 459-81, 1971.
2. Yule, G. U., The Statistical Study of Literary Vocabulary, Cambridge, 1944.
3. Massey, F. J., jr., "A Note on the Estimation of a Distribution Function by Confidence Limits," Annals of Mathematical Statistics, v. 21, pp 116-19, 1950.
4. Birnbaum, Z. W., "Numerical Tabulation of the Distribution of Kolmogorov's Statistic for Finite Sample Size," Journal of the American Statistical Association, v. 47, pp 425-41, September 1952.
5. Birnbaum, Z. W. and Tingey, F. H., "One-Sided Confidence Contours for Probability Distribution Functions," Annals of Mathematical Statistics, v. 22, pp 592-96, 1951.
6. Miller, I. H., "Table of Percentage Points of Kolmogorov Statistic," Journal of the American Statistical Association, v. 51, pp 111-21, March 1956.
7. Smirnov, N., "Tables for Estimating the Goodness of Fit of Empirical Distributions," Annal of Mathematical Statistics, v. 19, pp 279-81, 1948.
8. Lapin, L. L., Statistics for Modern Business Decisions, Jovanovich, 1973.
9. Kelley, T. L., The Kelley Statistical Tables, p. 3, Macmillan, 1938.
10. Altenbernd, L. and Lewis, L. L., Introduction to Literature: Stories, 2nd ed., Macmillan, 1970.
11. Barrett, L. L., Five Centuries of Spanish Literature, Dodd, Mead, & Co., 1963.
12. Medina, Jose Ramon, Antologia Venezolana, Editorial Gredos, 1962.
13. Instituto Internacional de Literatura Iberoamericana, An Anthology of Spanish American Literature, 2nd ed., Appleton-Century-Crofts, 1946.
14. Stallman, R. W. and Watter, R. E., The Creative Reader, 2nd ed., Ronald Press, 1962.

15. Saturday Evening Post, the, Best Modern Short Stories, Curtis
Books, 1965.